

(19) 日本国特許庁 (J P)

(12) 公表特許公報 (A)

(11) 特許出願公表番号
特表2001-505402
(P2001-505402A)

(43) 公表日 平成13年4月24日 (2001. 4. 24)

(51) Int.Cl. ⁷	識別記号	F I	テームト* (参考)
C 1 2 N 15/09	Z N A	C 1 2 N 15/00	Z N A A
A 0 1 K 67/027		A 0 1 K 67/027	
C 0 7 K 14/47		C 0 7 K 14/47	
16/18		16/18	
C 1 2 N 1/21		C 1 2 N 1/21	

審査請求 未請求 予備審査請求 有 (全 62 頁) 最終頁に続く

(21) 出願番号 特願平8-533715
(86) (22) 出願日 平成8年5月2日 (1996. 5. 2)
(85) 翻訳文提出日 平成9年11月10日 (1997. 11. 10)
(86) 国際出願番号 PCT/EP 96/01818
(87) 国際公開番号 WO 96/35784
(87) 国際公開日 平成8年11月14日 (1996. 11. 14)
(31) 優先権主張番号 195 16 776. 7
(32) 優先日 平成7年5月10日 (1995. 5. 10)
(33) 優先権主張国 ドイツ (DE)
(81) 指定国 EP (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, L U, MC, NL, PT, SE), CA, JP, MX, U S

(71) 出願人 ベーリンガー インゲルハイム インター
ナショナル ゲゼルシャフト ミット ベ
シュレンクテル ハフツング
ドイツ連邦共和国 デー55216 インゲル
ハイム アム ライン ポストファッハ
200
(72) 発明者 イェーヌヴァイン トーマス
オーストリア アー1030 ウィーン パー
リッヒガッセ 21-27
(72) 発明者 ライブレ ゲーツ
オーストリア アー1150 ウィーン コシ
ュタガッセ 9-13
(74) 代理人 弁理士 中村 稔 (外7名)

(54) 【発明の名称】 クロマチン調節遺伝子

(57) 【要約】

本発明はSETドメインを有するクロマチン調節遺伝子の
脱調節に関するものであり、このような脱調節は或る種
の癌症状に重要である。これらの遺伝子、特にSETドメ
インそのものはこのような症状の診断及び治療に使用し
得る。

【特許請求の範囲】

1. SETドメインを有するクロマチン調節タンパク質をコードする配列またはその部分配列を含むDNA分子であって、それらがEZH2、SUV39HまたはEZHI（これらの縮重変異体及び突然変異体を含む）をコードする図6、図7または図8に示されたヌクレオチド配列を含むことを特徴とするDNA分子。
2. DNA分子がcDNAであることを特徴とする請求の範囲第1項に記載のDNA分子。
3. DNA分子がヒト起源のものであることを特徴とする請求の範囲第1項または第2項に記載のDNA分子。
4. EZH2として知られている請求の範囲第3項に記載のcDNA。
5. SUV39Hとして知られている請求の範囲第3項に記載のcDNA。
6. DNA分子がEZH2のSETドメイン、またはその部分をコードする領域を含むことを特徴とする請求の範囲第4項に記載のDNA分子。
7. DNA分子がSUV39HのSETドメイン、またはその部分をコードする領域を含むことを特徴とする請求の範囲第5項に記載のDNA分子。
8. DNA分子がEZH2の優性-陰性突然変異体をコードすることを特徴とする請求の範囲第4項に記載のDNA分子。
9. DNA分子がSUV39Hの優性-陰性突然変異体をコードすることを特徴とする請求の範囲第4項に記載のDNA分子。
10. 原核宿主生物または真核宿主生物中の発現のために、発現調節配列に機能的に連結された、請求の範囲第2項に記載のcDNAを含む組換えDNA分子。
11. 請求の範囲第10項に記載の組換えDNAで形質転換された原核宿主生物または真核宿主生物。
12. 請求の範囲第4項に記載のcDNAの発現により得られる、組換えヒトクロマチン調節タンパク質EZH2またはそのフラグメント。
13. 請求の範囲第5項に記載のcDNAの発現により得られる、組換えヒトクロマチン調節タンパク質SUV39Hまたはそのフラグメント。
14. EZH2の抗体。

15. SUV39Hの抗体。
16. 請求の範囲第1項に記載のDNA分子の部分配列に相補性を有するアンチセンス（デオキシ）リボヌクレオチド。
17. SETドメインを有するクロマチン調節遺伝子の脱調節に起因し得るヒトの疾患の治療及び診断のためのクロマチン調節遺伝子のSETドメイン、またはその部分をコードするDNA分子。
18. SETドメインを有するクロマチン調節遺伝子の脱調節に起因し得るヒトの疾患の治療及び診断のための請求の範囲第6項または第7項に記載のDNA分子。
19. SETドメインを有するクロマチン調節遺伝子の脱調節に起因し得るヒトの疾患の治療及び診断のための請求の範囲第14項または第15項に記載の抗体。
20. SETドメイン、またはこのようなタンパク質の突然変異バージョンを有するクロマチン調節遺伝子の発現のためのトランスジーンを含むトランスジェニックマウス。
21. Ezh1及びSuv39hの内在性マウス遺伝子座が相同組換えにより中断される胚幹細胞から得られるノックアウトマウス。
22. SETドメイン、またはその突然変異バージョンを有する哺乳類のクロマチン調節遺伝子の同定方法であって、哺乳類のcDNAライブラリーまたはゲノムDNAライブラリーを非ストリンジェンシー条件下でSETドメインまたはその部分をコードするDNA分子でハイブリッドを形成することを特徴とする同定方法。

【発明の詳細な説明】

クロマチン調節遺伝子

本発明はクロマチンの構造及び機能の調節に役割を果たす遺伝子、並びに治療及び診断におけるそれらの使用に関する。

動原体、テロメア中並びに真正染色質性領域及び異質染色質性領域中の真核染色体の機能的組織化は、夫々の細胞分裂に関する遺伝情報の正確な複製及び分布を確実にするのに重要なメカニズムである。対照的に、腫瘍細胞は染色体の再配列、トランスロケーション及び異数性をしばしば特徴としている(Solomonら, 1991; Pardue, 1991)。腫瘍細胞中の増大された染色体不安定性をもたらすメカニズムは未だ明瞭にされていなかったが、ごく最近、酵母中のテロメアの位置効果(Renauldら, 1993; Buck及びShore, 1995; Allshireら, 1994)で始まって、ショウジョウバエ属中の班入り型位置効果(PEV) (Reuter及びSpierer, 1992)を経由して、ヒト白血病におけるトランスロケーション破損点の分析(Solomonら, 1991; Cleary, 1991)に至る幾つかの実験系が、脱調節された(deregulated)増殖に関与する幾つかの染色体タンパク質を同定することを可能にしていた。

第一に、SIR4-タンパク質の短縮バージョンの過剰発現(overexpression)が酵母の長い寿命をもたらすことがわかった(Kennedyら, 1995)。SIRタンパク質は正常交配型遺伝子座及びテロメアにおける多量体複合体の生成に寄与するので、過剰発現されたSIR4はこれらの異質クロマチン様複合体と干渉し、最終的に調節されない増殖をもたらし得るであろう。この仮定はヒトの癌の殆どの型における脱調節されたテロメア長さの発生の頻度と一致する(Counterら, 1992)。

第二に、ショウジョウバエ属中のPEVの遺伝子分析が、異質染色質性位置で、またホメオティック遺伝子クラスター内でクロマチンの構造を変化する幾つかの遺伝子産物を同定した(Reuter及びSpierer, 1992)。これらの遺伝子の幾つか、例えば、モジュロ(modulo) (Garzinoら, 1992)及びポリホメオティック(Smouse及びPerrimon, 1990)の突然変異がショウジョウバエ属の脱調節された細胞増殖または細胞死滅を生じ得る。第三に、ホメオティックショウジョウバエ属セレクター遺伝子のクロマチン構造のアクチベーター(トリソラックス(trithorax)ま

たはtrx-グループ)そしてまたリプレッサー(例えば、ポリコームまたはPc-グループ)の両方の哺乳類同族体が記載されていた。これらの中で、ヒトHRX/ALL-1(trx-グループ)がトランスロケーションにより誘発された白血病発生に関与することが示され(Tkachukら, 1992;Guら, 1992)、またマウスbmi(Pc-グループ)の過剰発現がリンパ腫の形成をもたらすことが示されていた(Hauptら, 1991;Bunkら, 1991;Alkemaら, 1995)。染色体タンパク質の機能に関するモデルは、それらが複合体中のアクチベーターとリプレッサーの間のバランスに応じて周囲のクロマチン領域の縮合度を決定する多量体複合体を生成することを結論させる(Lockeら, 1988)。複合体の成分の一つの過剰発現により生じたこの平衡のシフトは真正染色質性領域及び異質染色質性領域の新しい分布を示した(Buck及びShore, 1995; Reuter及びSpierer, 1992; Eissenbergら, 1992)。この用量効果は所定の遺伝子座におけるクロマチン構造を不安定にすることができ、これが最後の分析において正常状態から形質転換状態への転移をもたらす。

クロマチン構造を変化することができるプロトオンコジーンとしてのHRX/ALL-1及びbmiの特性決定にもかかわらず、クロマチンと相互作用する哺乳類遺伝子産物に関する本発明者らの知識は依然として非常に制限される。対照的に、ショウジョウバエ属中のPEVの遺伝子分析により、クロマチンレギュレーターに関する約120の対立遺伝子が記載されていた(Reuter及びSpierer, 1992)。最近、カルボキシ末端領域が、陽性(t r x、trx-グループ)及び陰性(E(z)、Pc-グループ)ショウジョウバエ属クロマチンレギュレーターに共通である配列の類似性により同定された(Jones及びGelbart, 1993)。更に、このカルボキシ末端はまたショウジョウバエ属中のクロマチン分布の優性サプレッサーであるS_u(var)3-9に保存される(Tschierschら, 1994)。

本発明は、“SET”と称されるこのタンパク質ドメイン(Tschierschら, 1994)がそれらの進化的保存及び拮抗遺伝子産物中のそれらの存在のためにそれらの発育ヒストリーに関して重要である哺乳類クロマチンレギュレーターの新しい遺伝子ファミリーを形成するという前提から開始した。更に、HRX/ALL-1は別にして、SETドメイン遺伝子のグループのその他の員の特性決定は、悪性形質転換をおそらくもたらすクロマチンの構造変化の原因であるメカニズムを説明することを

助

ける。

それ故、本発明の目的はヒトクロマチン調節遺伝子を同定し、それらの機能を明らかにし、それらを診断及び治療に使用することであつた。

この目的を達成するために、最初にSETドメインの配列情報がヒトcDNAバンクからショウジョウバエ属のSETドメイン遺伝子に相同のヒトcDNAを得るのに使用された。E(z)及びSu(var)3-9のヒト同族体を構成する2種のcDNAが得られた。相当するヒト遺伝子がEZH2及びSUV39H（下記を参照のこと）と称された。加えて、EZH2の変異体形態が同定され、これがEZH1と称された。

こうして、本発明は、DNA分子が図6または図7に示されたヌクレオチド配列を有することを特徴とする、SETドメインを有するクロマチン調節タンパク質をコードする配列、またはその部分配列を含むDNA分子に関する。本発明のDNA分子は以下“本発明の遺伝子”とも称される。

本発明の遺伝子はEZH2及びSUV39Hと称される。それらは最初に“HEZ-2”及び“H3-9”として知られていた。

別の局面によれば、本発明はこれらの遺伝子に由来するcDNA（その縮重変異体；機能性クロマチンレギュレーターをコードする突然変異体及び遺伝子重複にトレースバックし得る変異体を含む）に関する。この例がEZH1であり、その部分配列が図8中にEZH2と比較して示される。

本発明の範囲に示された目的を達成するために、以下の特別な操作に従った。保存されたSETドメインの配列情報から開始して、ヒトB細胞特異性cDNAライブラリーが、低下されたストリンジェンシーのもとに、E(z)及びSu(var)3-9のSETドメインをコードする混合ショウジョウバエ属DNAプローブでスクリーニングされた。500,000のブランクから、40の一次ファージが選択された。スクリーニングの別の2ラウンド後に、31種のファージが真正E(z)配列をコードし、5種のファージがE(z)変異体を構成することが明らかになった。対照的に、2種のファージのみがSu(var)3-9単独のSETドメインを含むプローブとハイブリッドを形成した。

ファージインサートがポリメラーゼ連鎖反応(PCR)により増幅され、制限マッピング及び部分配列決定により分析された。代表的なcDNAインサートがサブ

クローン化され、それらの全長にわたって配列決定された。5'末端が陽性ファージを5'-DNAプローブでもう一度スクリーニングすることにより単離され、その後、サブクローニング後に、完全cDNAが得られた。

E(z)のヒト同族体をコードする完全cDNAがEZH2と称され、Su(var)3-9のヒト同族体をコードするDNAがSUV39Hと称された。全体で、ショウジョウバエ属タンパク質とヒトタンパク質の間のアミノ酸の同一性はEZH2について61%の量であり、SUV39Hについて43%の量であり、一方、C末端SETドメインが非常に高度に保存される(EZH2について88%そしてSUV39Hについて53%)。配列比較は相同性のその他の明らかな領域、例えば、EZH2中のシステインに富むドメイン及びSUV39H中のクロモボックス(Chromo-Box)を示した(ポリコーム中で、クロモボックスはDNAとクロマチンの間の相互作用に必須のドメインであることが示された; Messmerら, 1992)。対照的に、ショウジョウバエ属タンパク質のアミノ末端GTP結合モチーフを含む207アミノ酸はヒト同族体SUV39Hから不在である。

ショウジョウバエ属遺伝子とヒト遺伝子の間のアミノ酸配列の比較が図1及び2に示される。

図1はEZH2とゼスト(zeste)のショウジョウバエ属エンハンサー(E(z))の間のアミノ酸配列の比較を示す。保存されたカルボキシ末端SETドメイン(陰影付きのボックス)及びCysに富む領域(Cys基が強調されている)の同一アミノ酸が示される。推定の核配置シグナルが下線を施される。

図2はヒト同族体SUV39Hとショウジョウバエ属Su(var)3-9の間のアミノ酸配列の比較を示す。保存されたカルボキシ末端SETドメイン(陰影付きのボックス)及びクロモドメイン(更に暗い陰影付きのボックス)の同一アミノ酸が示される。推定の核配置シグナルが下線を施される。図面の上部に2種のタンパク質構造の図式要約があり、これはヒト同族体中で207アミノ酸がN末端で欠いていることを示す。

翻訳コンセンサス配列がまたヒトSUV39H-cDNAの開始ATGの環境中、更にはSu(v

ar)3-9中の相当する内部位置に存在するので、ショウジョウバエ属タンパク質は付加的なエクソンを含むべきであり、これらは進化のその後の段階で機能に必要なではなくなった。

この仮説の正確なことはショウジョウバエ属中で5'末端で完全であり、または短縮されたヒトSUV39H-cDNA及びSu(var)3-9のcDNAを発現することにより確認し得る。更に、MG-44として知られている別のcDNAが記載され(以下を参照のこと)、これはまたショウジョウバエ属遺伝子の5'末端に欠いている。SUV39HのヒトcDNAの他に、相同遺伝子座がまたマウスで単離され(Suv39h;以下を参照のこと)、その配列分析及びプロモーター構造がショウジョウバエ属Su(var)3-9と較べて哺乳類相同遺伝子のアミノ末端短縮を明らかに確認する。

本発明の範囲内で行われたDNAプロット分析は、Su(var)3-9の哺乳類相同遺伝子が個々の遺伝子座によりマウス及びヒト中で提示され、一方、E(z)の哺乳類相同遺伝子がマウス及びヒト中で二つの別個の遺伝子座によりコードされることを示す。また、第二のヒト遺伝子座(EZH1として知られている)がヒトcDNAライブラリーから単離されたクローンの大半とはそれらの3'隣接配列を異にする少数のcDNA変異体を特性決定することにより確認された。配列決定された領域中のEZH2とEZH1の相違が図8に示される。このEZH1のSETドメインはEZH2と較べて突然変異を示す。更に、本発明者らが単離したEZH1変異体(おそらく、異常にスプライシングされたcDNA)は47C末端アミノ酸だけタンパク質を短縮する読み取り枠中に配置された終止コドンを含む。図8は上部ライン中の位置1844~2330からのEZH2cDNAのヌクレオチド配列、5'スプライシング部位及び下線を施されている潜在的な終止コドンを示す。EZH1変異体のcDNAの部分配列をEZH2配列に起因するために、本発明者らはウィスコンシンGCGネットワークサービスのギャッププログラムを使用した。EZH1中の早期終止コドン(位置353)が下線を施されている。保存されたSETドメインをコードする配列が目立つようにされる。更に、異常な転写産物B52(以下を参照のこと)の3'末端(EZH1中の位置151)が示される。利用可能な配列にわたって、B52はEZH1に97%同一であり、かつEZH2に72%同一であることがわかった。EZH1とEZH2との配列比較並びにヒト及びマ

ウス中に二つの別個のE(z)相同遺伝子座があるという知見は、遺伝子重複が哺乳類中で起こったことを結論させる。

ジーンバンクデータベース中のcDNA配列と比較して、腫瘍組織中の異常な転写産物に由来するデータベースに記録された或るcDNA部分配列は本発明の

cDNAの突然変異バージョンを構成することが驚くことにわかった。

一方、乳癌及び卵巣癌にかかりやすくする遺伝子であるBRCA1に関する研究において、271ヌクレオチドを有する部分cDNA配列が単離され、B52として知られ、これはSETドメインの突然変異体をコードし、それがヒト染色体17q21にマッピングされた(Friedmanら, 1994)。本発明の範囲内で、B52は本発明のEZH1-cDNA変異体(上記を参照のこと)と97%の同一性を示すことが驚くことにわかった。EZH1はおそらく再活性化が脱調節された増殖に役割を果たす遺伝子であるかもしれない。

一方、cDNA(2,800ヌクレオチド; MG-44)がヒト染色体Xp11(Geraghtyら, 1993)から単離され、その領域は網膜及び滑膜の肉腫の変性疾患にかかりやすくする。このcDNAは本発明のSUV39H-cDNAと98%の同一性を有することが驚くことにわかった。

こうして、本発明の範囲内で調製された新規遺伝子は、或る種の癌とクロマチンレギュレーター中の突然変異の間の相関関係を推定することを可能にする。MG-44-cDNAの場合、それがクロモドメイン及びSETドメインを中断する多数の点突然変異及びフレームシフト突然変異を有するので、本発明のSUV39H-cDNAを使用してSu(var)3-9とMG-44の相関関係を明らかにすることが初めて可能になった。

既に記載された配列は別にして、ジーンバンク配列データベースはまた、SETタンパク質ファミリーの別のヒト員としてショウジョウバエ属

突然変異バージョンが従来知られていないSETドメインを有する現在唯一のヒト遺伝子であるが、KG-1はSETドメインをアミノ末端ハーフ及びカルボキシ末端ハーフに開裂する342アミノ酸の挿入を有する。おそらく、このKG-1-cDNAは異常にスプライシングされた変異体を構成する。何となれば、その挿入の両末端に5'及び3'コンセンサススプライシング部位があるからである。

合計で、SETタンパク質ファミリーの5種の現在知られているヒト員のうちの4種が変化を受けており、その全ての変化がSETドメインを突然変異する(HRX/ALL-1、EZH1/B52、SUV39H/MG-44及びKG-1)。更に、三つの場合に、トランスロケーション破損点または不安定な染色体領域の付近の相当するヒト遺伝子座がマッピングされた(HRX/ALL-1、EZH1/B52及びSUV39H/MG-44)。図3はヒトSETドメイン遺伝子の異常な転写産物を示す。図の左に、適当な染色体に関する5種の現在知られているSETドメイン遺伝子の位置がある。図は、とりわけ、異常なcDNAがトランスロケーション破損点または不安定なクロマチン領域にマッピングされた3種の遺伝子(HRX/ALL-1、EZH1/B52及びSUV39H/MG-44)を示す。示された5種のSETドメイン遺伝子のうちの4種が突然変異を有し、その全てが図中に暗いボックスにより示されているカルボキシ末端SETドメインを中断する。トランスロケーションはHRXのアミノ末端ハーフを無関係の遺伝子配列（これはENLと称される点付きのボックスとして示される）に連結する。突然変異及び早期終止コドンはEZH1/B52のSETドメインを変化する。点突然変異及びフレームシフト突然変異はMG44中のクロモドメイン及びSETドメインを中断する。大きい挿入はKG-1のSETドメインを二つのハーフに開裂する。現在、G9aについて異常な転写産物は知られていない。B52中のシステインに富むクラスターが点付きのボックスとして示される。HRX/ALL-1では、メチルトランスフェラーゼとの相同性の領域が陰影付きのボックスとして示され、A/Tフックが垂直の線として示される。夫々の場合の真正遺伝子の名称が図の右手側に示される。

SETタンパク質ファミリーの哺乳類遺伝子であるHRX/ALL-1がトランスロケーション誘導白血病発生と関連していたという事実(Tkachukら, 1992;Guら, 1992)は、SETドメインを含むタンパク質が遺伝子発現のクロマチン依存性変化を同時決

定する発育の重要なレギュレーターであるだけでなく、突然変異後に、それらがまた正常な細胞増殖を中断するという強力な指示である。

従来記載された突然変異の全てがSETドメインの一次構造を中断するので、正常な状態から形質転換された状態への転移に重要な役割を果たすのはSETドメインそのものであることを仮定することは正当である。更に、SETドメインは酵母からヒトに至るまで生じる遺伝子産物中のその進化的保存に鑑みて重要な役割を

有するものと考えられる。

図4はSETドメインタンパク質の進化的保存を示す。ウィスコンシンGCGネットワークサービスのtfastaプログラムを使用して、SETドメインに対する相同性を有するタンパク質及び読み取り枠が同定された。図は酵母からヒトに至るまでの代表的な選択を示す。番号はアミノ酸を示す。カルボキシ末端SETドメインが黒色のボックスにより表され、Cysに富む領域が暗い点付きのボックスにより示され、Su(var)3-9中のGTP結合モチーフが明るい点付きのボックスにより示され、またSu(var)3-9及びH3-9のクロモドメインが明るい点付きのオープンボックスにより示される。メチルトランスフェラーゼに相同である領域(trx及びHRX)が陰影付きのボックスとして示される。A/Tフックが垂直の線により示される。また、別のSerに富む領域(C26E6.10中のS)及びGluに富む領域(G9a中のE)またはアンキリンリピート(G9a中のANK)が強調される。YHR119(ジーンバンク受理番号U00059)及びC26E.10(ジーンバンク受理番号U13875)は機能上の特性決定をしないでデータバンクに最近入れられたコスミドの読み取り枠である。％はヒト及びショウジョウバエ属のタンパク質間の合計のアミノ酸同一性を示す。

図5はSETドメイン中のアミノ酸の間の一致を示す。図4に示された遺伝子のSETドメインは、ウィスコンシンGCGネットワークサービスのパイルアップ・プログラムを使用して配置された。KG-1SETドメインを比較するために、SETドメインを二つのハーフに分ける大きいアミノ酸インサートがパイルアップの前に除去された。10の一致のうちの8を有するアミノ酸位置が強調される。

本発明の範囲内で定められた基準に基づいて、SETドメインを有する遺伝子は脱調節された増殖のクロマチン依存性発生に関与することが明らかになる。こうし

て、これらの遺伝子またはそれらに由来するcDNA並びに部分配列及び突然変異配列がこのような増殖に起因し得る疾患の治療及び診断に使用し得る。

正常細胞及び形質転換細胞の間のSETドメインRNAの転写レベルの差異は、SETドメイン遺伝子の発現が脱調節される疾患の診断パラメーターとして使用し得る。

こうして、SETドメインそのものまたはその部分をコードするオリゴヌクレオチドは、SETドメインが突然変異される癌の或る型を診断するために診断マーカ

ーとして使用し得る。本発明の範囲内のSETドメイン遺伝子配列の診断上の使用に関する本発明のcDNAまたはその部分の機能の詳細な分析について、相同のマウスcDNAがEZH1 (Ezh1) 及びSUV39H (Suv39h) から単離された。正常なマウス発育中のEzh1遺伝子活性を研究するために“RNase保護”分析においてSETドメインをコードするマウス特異性DNAプローブを使用する場合、bmiの発現プロフィール(Hauptら, 1991)に似ている若干広い発現プロフィールが明らかになった。そのマウス配列を用いて行われた分析は、種々のヒト細胞培養系における未熟前駆細胞、腫瘍細胞及び分化細胞の間のRNAの量を比較するためにヒト配列で拡大された。それ故、SETドメインが特別な癌の診断腫瘍マーカーまたは一般の診断パラメーターとして適しているかを見出すために、関連実験マニュアル(Sambrook, J., Fritsch, E. F. 及びManiatis, T., 1989, Cold Spring Harbor Laboratory Press)に記載されたようにRNA濃度を測定するための現行の方法、例えば、ノーザンブロット、S1-ヌクレアーゼ保護分析またはRNase保護分析を使用することが可能である。

SETドメインが特別な突然変異を受ける頻度を調べるために、SET特異性DNAプローブを使用して一本鎖コンホメーション多形性を分析することが可能である(SSCP; Gibbonsら, 1995)。

SET特異性DNAプローブが診断マーカーとして使用し得る癌の型は乳癌(EZH1; Friedmanら, 1994)、滑膜癌(SUV39H; Geraghtyら, 1993)及び白血病である。

SETドメイン遺伝子のヌクレオチド配列の知識に鑑みて、cDNA配列に由来する相当するタンパク質(これらはまた本発明の目的である)を、それらをコー

ドするcDNAを好適なベクターに挿入し、それらを宿主生物中で発現することにより、組換え形態で生成することが可能である。組換えタンパク質を生成するのに使用される技術は当業者に公知であり、また関連マニュアル(Sambrook, J., Fritsch, E.F.及びManiatis, T., 1989, Cold Spring Harbor Laboratory Press)から採用されてもよい。

こうして、本発明は、別の局面において、EZH2、SUV39HまたはEZH1をコードするDNA及びそれらに機能的に連結する発現調節配列を含む組換えDNA分子、

並びにそれにより形質転換された宿主生物に関する。

本発明の組換えタンパク質はクロマチンまたは異質クロマチン複合体のその他の員とのSETドメインタンパク質の相互作用を分析するのに使用し得る。これらの複合体の活性の様式に関してこうして得られた知見から開始して、それに関するメカニズムにおける標的された介入に関する詳細な可能性が特定され、治療用途に使用し得る。

SETドメインの機能を更に分析するのに利用できる研究が、例えば、ヒトEZH2またはSUV39Hをコードし、かつエピトープ（それに対する抗体が利用できる）を備えているcDNAをin vitro及び組織培養で発現することにより行われる。適当なエピトープ特異性抗体による免疫沈殿後に、EZH2及びSUV39Hがin vitroで互いに相互作用することができること、また複合体形成がEZH2及び／またはSUV39Hとその他のクロマチンレギュレーターの間でin vivoで起こることを証明することが可能である。

異質クロマチンタンパク質の種々の員の間複合体形成がそれらの機能に必須であることがその他の著者ら(DeCamillisら, 1992; Rastelliら, 1993; Orland及びParo, 1993)により既に推定されていた。本発明のSETドメイン遺伝子の利用可能性に鑑みて、SET領域が相互作用のために機能するドメインを構成すること、またはそれが多量体の異質染色質性複合体の形成に寄与することを測定することが可能である。同様に、SETドメインがGAGA因子(Adamsら, 1992)を含む種々のクロマチンレギュレーターのアミノ末端BTBドメインと同様の抑制機能を有することを測定することが可能である。全部で、エピトープを備えたEZH2及びSUV39Hタ

タンパク質との相互作用の分析はSETドメインの機能の更なる特性決定を可能にする。これは、例えば、遺伝子治療方法を使用してSETドメインcDNA配列の優性-陰性変異体を細胞に導入することにより、脱調節された活性に対して作用することの可能性を広げる。このような変異体は、例えば、最初にSETタンパク質の機能性ドメイン、例えば、DNA/クロマチン相互作用またはタンパク質間の相互作用の原因の配列部分を特定し、次いで無傷の機能性タンパク質により生じる脱調節された増殖と競合するために、一種以上の関連ドメインまたはその部分により短縮されたDNA配列を当該細胞中で発現することにより得られ

る。

また、本発明のcDNAの利用可能性は、トランスジェニック動物、例えば、マウスを生産することを可能にし、これらの動物中で、SETドメイン遺伝子が過剰発現でき（“機能の獲得”）、またはこれらの遺伝子がスイッチオフし得る（“機能の損失”）。本発明の遺伝子の相当する動物配列、特にマウス配列が最終分析に使用される。また、これらのマウスは本発明の目的である。

特に、本発明の遺伝子の対立遺伝子がマウスに導入される“機能の獲得”分析は、腫瘍形成のクロマチン依存性要件におけるEZH2及びSUV39Hの原因となる関与について最終の結論を与える。“機能の獲得”分析について、ヒトEZH2及びSUV39Hの完全cDNA配列並びにこれらの突然変異バージョン、例えば、EZH1/B52及びBMG-44が高い発現率を可能にするベクター、例えば、ヒト β -アクチンプロモーターを有するプラスミド、及び免疫グロブリン(E μ)のH鎖のエンハンサーそしてまたモロニーウイルスエンハンサー(Mo-LTR)により誘導し得る。最近、bmi遺伝子（これはEZH2と共通に陰性クロマチンレギュレーターのPcグループに属する）のE μ /Mo-LTR依存性過剰発現がトランスジェニックマウス中にリンパ腫を生じるのに充分であることが示されていた(Alkemaら, 1995)。

“機能の損失”分析において、Ezh1及びSuv39hの内在性マウス遺伝子座が胚幹細胞中で相同組換えにより中断されるものと仮定すると、in vivo遺伝子機能の損失がマウスの異常な発育をもたらすことを測定することが可能である。

これらのin vivo系の結果として、EZH2及びSUV39Hの活性が確認し得る。これ

らの系はまたヒト遺伝子治療と関連する動物モデルの基礎を形成する。

本発明のDNA配列またはそれから誘導された配列（例えば、相補アンチセンスオリゴヌクレオチド）は、機能性遺伝子配列を導入することにより、例えば、アンチセンスオリゴヌクレオチドを使用して、遺伝子発現を抑制することにより、または優性-陰性突然変異体をコードする配列を導入することにより、治療すべき疾患が機能性遺伝子配列の不在の結果として、または相当する遺伝子の過剰発現の結果としてクロマチンの脱調節に起因し得ることに応じて、遺伝子治療に使用し得る。当該DNA配列は高等真核生物細胞のトランスフェクションのための通常の方法（これはウイルスベクター（レトロウイルス、アデノウイルス、アデノ

ノ関連ウイルス）を使用する遺伝子移入を含んでもよい）を使用して、またはレセプター介在性エンドサイトーシスに基く非ウイルス系を使用して細胞に挿入し得る。普通の方法の概説が、例えば、Mitani及びCaskey, 1993; Jolly, 1994; Vile及びRussell, 1994; Tepper及びMule, 1994; Zatloukalら, 1993; W093/07283により与えられる。

本発明の遺伝子の発現を抑制するために、転写の機構に干渉する低分子物質を使用することがまた可能である。遺伝子の5'調節領域を分析した後に、例えば、W092/13092に記載された方法を使用して、この領域との関連転写因子の相互作用を完全または部分的に阻止する物質についてスクリーニングすることが可能である。

脱調節された増殖の抑制はまたEZH2タンパク質またはSUV39Hタンパク質の相当する抗体、例えば、ヒト抗体またはヒト化抗体を治療上使用することにより遺伝子産物に作用し得る。このような抗体は既知の方法により、例えば、Malavsi及びAlbertini, 1992またはRhein, 1993により記載されたようにして産生される。

また、本発明は治療上または診断上使用し得るEZH2またはSUV39Hの抗体に関する。

図面の要約

図1：EZH2とE(z)の間のアミノ酸配列の比較

- 図2: SUV39HとSu(var)3-9の間のアミノ酸配列の比較
- 図3: ヒトSETドメイン遺伝子の異常な転写産物
- 図4: SETドメインタンパク質の進化的保存
- 図5: SETドメイン中のアミノ酸の一致
- 図6: EZH2のDNA及びアミノ酸配列
- 図7: SUV39HのDNA及びアミノ酸配列
- 図8: ヒトEZH2及びEZH1のcDNAの間の部分配列の比較

実施例

a) cDNAライブラリーの調製

Bardwell及びTreisman, 1994により記載されたようなヒトB細胞特異性cDNAライブラリーを、ポリ(A)⁺-RNAをヒトBJA-B細胞から単離し、それをポリ(dT)₁₅プライミングにより逆転写酵し、それを二本鎖cDNAに変換することにより調製した。配列5' AATTCTCGAGCTCGTCGACAのEcoRIアダプターの添加後に、cDNAをバクテリオファージgt10のEcoRI部位につないだ。ライブラリーの増殖及び増幅をE. coli C600中で行った。

b) DNAプローブの調製

E(z)及びSu(var)3-9の保存SETドメインをコードするショウジョウバエ属DNAプローブをポリメラーゼ連鎖反応(PCR)により公表されたショウジョウバエ属配列(Jones及びGelbart, 1993; Tschierschら, 1994)に基いて調製した。ショウジョウバエ属メラノガスター(*Drosophila melanogaster*) DNA (クロンテック) 1 µgを2種のプライマーE(z) 1910(5' ACTGAATTCGGCTGGGGCATCTTTCTTAAGG)及びE(z) 2280(5' ACTCTAGACAATTCCATTTTCAGCTCTATG)とともにPCR増幅(94℃で30秒、55℃30秒及び72℃で30秒の35サイクル)にかけた。Su(var)3-9に関する相当するSETドメインプローブを、同サイクル条件を使用して、プライマーsuvar.up(5' ATATAGTACTTCAAGTCCATTCAAAGAGG)及びsuvar.dn(5' CCAGGTACCGITGGTCTGTTT AAGACCG)の対を用いてプラスミドDNA(Tschierschら, 1994; クローンM4) 10ngから増幅した。得られたSETドメインDNAフラグメントをゲル精製し、増幅された配列の正確さを確かめるために部分配列決定した。

c) cDNAライブラリーのスクリーニング

5 x 10⁵ プラーク形成単位 (pfu) を *E. coli* C600 のバクテリア宿主株の培養液 (10 mM の MgSO₄ 中 0.5 の光学密度 OD₆₀₀ で懸濁した) 5 ml とともに 37℃ で 15 分間 インキュベートし、ついで大きい (200 mm x 200 mm) 予熱された LB 皿に注入した。37℃ で一夜増殖した後、ファージをナイロン膜 (ジーンスクリーン) に吸収させた。吸収されたファージを含む面を上向きにして、膜を変性溶液 (1.5 M の NaCl、

0.5 M の NaOH) 中で 30 秒間浮遊させ、次いで変性溶液に 60 秒間浸漬し、最後に 3 M の NaCl、0.5 M の トリス pH8 中で 5 分間中和した。次いで膜を 3xSSC 中で素早くすすぎ、ファージ DNA を UV 架橋によりナイロンフィルターに固定した。フィルターをチャーチ緩衝液 (1% の BSA、1 mM の EDTA 及び 0.5 M の NaHPO₄、pH7.2) 30 ml 中で 30 分間にわたって 50℃ でプレハイブリダイゼーションし、次いで (E(z)-SET 及び Su(var)3-9-SET) の放射能標識した DNA プローブ混合物 2 x 10⁶ cpm を添加した。その DNA プローブをレジプライム (RediPrime) キット (アメーシヤム) を使用してランダムプライミングにより調製した。ハイブリダイゼーションを 50℃ で一夜行なった。ハイブリダイゼーション溶液を除去した後、フィルターを周囲温度で 2xSSC、1% の SDS 中で 10 秒間洗浄し、次いで 50℃ で 10 秒間洗浄した。フィルターをサララップで包み、増感フィルムを使用してオートラジオグラフィーにかけた。

オートラジオグラムを整合することにより、陽性ファージコロニーを最初のプレートで同定し、パスツールピペットの大きな端部を使用して、相当する寒天フラグメントを除去した。ファージプールを、二三滴の CHCl₃ を含む、SM 緩衝液 (H₂O 1 リットル中、NaCl 5.8 g、MgSO₄·H₂O 2 g、50 ml の トリス pH7.5、5 ml の 2% のゼラチン) 1 ml 中で 4℃ で一夜溶離した。個々の良く単離された陽性プラーク (第三ラウンドでプレート当たり 20~100 のプラーク) を得るために、ファージ溶解産物をスクリーニングの第二ラウンド及び第三ラウンドのために塗布した。

d) 配列分析

組換えファージからの cDNA を p ブルースクリプト KS (ストラタゲン) のポリリンカーにサブクローン化し、ジデオキシ方法を使用して自動配列決定装置 (アプライド・バイオシステムズ) 中で配列決定した。得られた遺伝子当たり少な

くとも2種の独立の単離物の完全配列をプライマーウォーキングにより決定した。配列をGCGソフトウェアパッケージ（ウィスコンシン大学）で分析し、“ブラスト・アンド・ファスタ”または“tfasta”ネットワークサービスを使用して相同性に関する研究を行った。EZH2及びSUV39Hの完全配列を図6及び7に示す。

Literature

- Adams et al., 1992, Genes & Dev. 6, 1589-1607.
- Alkema et al., 1995, Nature 374, 724-727.
- Allshire et al., 1994, Cell 76, 157-169.
- Bardwell and Treisman, 1994, Genes & Dev. 8, 1644-1677.
- Brunk et al., 1991, Nature 353, 351-355.
- Buck and Shore, 1995, Genes & Dev. 9, 370-384.
- Cleary, 1991, Cell 66, 619-622.
- Counter et al., 1992, Embo J. 11, 1921-1928.
- DeCamillis et al., 1992, Genes & Dev. 6, 223-232.
- Eissenberg et al., 1992, Genetics 131, 345-352.
- Friedman et al., 1994, Cancer Research 54, 6374-6382.
- Garzino et al., 1992, Embo J. 11, 4471-4479.
- Géraghty et al., 1993, Genomics 16, 440-446.
- Gibbons et al., 1995, Cell 80, 837-845.
- Gu et al., 1992, Cell 71, 701-708.
- Haupt et al., 1991, Cell 65, 753-763.
- Jolly, D., 1994, Cancer Gene Therapy 1, 51.
- Jones and Gelbart, 1993, MCB 13 (10), 6357-6366.
- Kennedy et al., 1995, Cell 80, 485-496.
- Locke et al., 1988, Genetics 120, 181-198.
- Malavsi, F. and Albertini, A., 1992, TIBTECH 10, 267-269.
- Messmer et al., 1992, Genes & Dev. 6, 1241-1254.
- Milner and Campbell, 1993, Biochem. J. 290, 811-818.
- Mitani, K. and Caskey, C.T., 1993, Trends in Biotechnology 11, 162-166.
- Nomura et al., 1994, Unpublished. GeneBank accession number: D31891.
- Orlando and Paro, 1993, Cell 75, 1187-1198.
- Pardue, 1991, Cell 66, 427-431.
- Rastelli et al., 1993 Embo J. 12, 1513-1522.
- Renauld et al., 1993, Genes & Dev. 7, 1133-1145.
- Reuter and Spierer, 1992, BioEssays 14, 605-612.
- Rhein, R., 1993, The Journal of NIH Res. 5, 40-46.
- Smouse and Perrimon, 1990, Dev. Biol. 139, 169-185.

- Solomon et al., 1991, Science 254, 1153-1160.
- Tepper, R.I. and Mule, J.J., 1994, Human Gene Therapy 5, 153.
- Tkachuk et al., 1992, Cell 71, 691-700.
- Tschiersch et al., 1994, Embo J. 13 (16), 3822-3831.
- Vile, R. and Russel S., 1994, Gene Therapy 1, 88.
- Zatloukal, K., Schmidt, W., Cotten, M., Wagner, E., Stingl, G. and Birnstiel, M.L., 1993, Gene 135, 199.

[配列表]

(2) 配列番号：1 の情報：

(i) 配列の特徴：

(A) 長さ：2600 塩基対

(B) 型：ヌクレオチド

(C) 鎖：一本鎖

(D) トポロジー：直鎖状

(ii) 分子の種類：cDNA to mRNA

(iii) ハイボセティカル：No

(iv) アンチセンス：No

(vi) 起源：

(A) 生物名：ヒト

(G) 細胞の種類：B-cell

(ix) 特徴：

(A) 名称／キー：5' UTR

(B) 存在位置：1..89

(ix) 特徴：

(A) 名称／キー：CDS

(B) 存在位置：90..2330

(ix) 特徴：

(A) 名称／キー：3' UTR

(B) 存在位置：2331..2600

(xi) 配列：配列番号：1：

AGGCACTGGA GCCCCGGCGG CGGCGGCGGC GCGCGCGGG GCGACGCGC GGGAACAACG 60

CGAGTCGGCG CGCGGGACGA AGAATAATC ATG GGC CAG ACT GGG AAG AAA TCT 113

Met Gly Gln Thr Gly Lys Lys Ser

1

5

GAG AAG GGA CCA GTT TGT TGG CGG AAG CGT GTA AAA TCA GAG TAC ATG	161
Glu Lys Gly Pro Val Cys Trp Arg Lys Arg Val Lys Ser Glu Tyr Met	
10 15 20	
CGA CTG AGA CAG CTC AAG AGG TTC AGA CGA GCT GAT GAA GTA AAG AGT	209
Arg Leu Arg Gln Leu Lys Arg Phe Arg Arg Ala Asp Glu Val Lys Ser	
25 30 35 40	
ATG TTT AGT TCC AAT CGT CAG AAA ATT TTG GAA AGA ACG GAA ATC TTA	257
Met Phe Ser Ser Asn Arg Gln Lys Ile Leu Glu Arg Thr Glu Ile Leu	
45 50 55	
AAC CAA GAA TGG AAA CAG CGA AGG ATA CAG CCT GTG CAC ATC CTG ACT	305
Asn Gln Glu Trp Lys Gln Arg Arg Ile Gln Pro Val His Ile Leu Thr	
60 65 70	
TCT GTG AGC TCA TTG CGC GGG ACT AGG GAG TGT TCG GTG ACC AGT GAC	353
Ser Val Ser Ser Leu Arg Gly Thr Arg Glu Cys Ser Val Thr Ser Asp	
75 80 85	
TTG GAT TTT CCA ACA CAA GTC ATC CCA TTA AAG ACT CTG AAT GCA GTT	401
Leu Asp Phe Pro Thr Gln Val Ile Pro Leu Lys Thr Leu Asn Ala Val	
90 95 100	
GCT TCA GTA CCC ATA ATG TAT TCT TGG TCT CCC CTA CAG CAG AAT TTT	449
Ala Ser Val Pro Ile Met Tyr Ser Trp Ser Pro Leu Gln Gln Asn Phe	
105 110 115 120	
ATG GTC GAA GAT GAA ACT GTT TTA CAT AAC ATT CCT TAT ATG GGA GAT	497

Met Val Glu Asp Glu Thr Val Leu His Asn Ile Pro Tyr Met Gly Asp	
125 130 135	
GAA GTT TTA GAT CAG GAT GGT ACT TTC ATT GAA GAA CTA ATA AAA AAT	545
Glu Val Leu Asp Gln Asp Gly Thr Phe Ile Glu Glu Leu Ile Lys Asn	
140 145 150	
TAT GAT GGG AAA GTA CAC GGG GAT AGA GAA TGT GGG TTT ATA AAT GAT	593
Tyr Asp Gly Lys Val His Gly Asp Arg Glu Cys Gly Phe Ile Asn Asp	
155 160 165	
GAA ATT TTT GTG GAG TTG GTG AAT GCC CTT GGT CAA TAT AAT GAT GAT	641
Glu Ile Phe Val Glu Leu Val Asn Ala Leu Gly Gln Tyr Asn Asp Asp	
170 175 180	
GAC GAT GAT GAT GAT GGA GAC GAT CCT GAA GAA AGA GAA GAA AAG CAG	689
Asp Asp Asp Asp Asp Gly Asp Asp Pro Glu Glu Arg Glu Glu Lys Gln	
185 190 195 200	
AAA GAT CTG GAG GAT CAC CGA GAT GAT AAA GAA AGC CGC CCA CCT CGG	737
Lys Asp Leu Glu Asp His Arg Asp Asp Lys Glu Ser Arg Pro Pro Arg	
205 210 215	
AAA TTT CCT TCT GAT AAA ATT TTT GAA GCC ATT TCC TCA ATG TTT CCA	785
Lys Phe Pro Ser Asp Lys Ile Phe Glu Ala Ile Ser Ser Met Phe Pro	
220 225 230	
GAT AAG GGC ACA GCA GAA GAA CTA AAG GAA AAA TAT AAA GAA CTC ACC	833
Asp Lys Gly Thr Ala Glu Glu Leu Lys Glu Lys Tyr Lys Glu Leu Thr	

235	240	245	
GAA CAG CAG CTC CCA GGC GCA CTT CCT CCT GAA TGT ACC CCC AAC ATA			881
Glu Gln Gln Leu Pro Gly Ala Leu Pro Pro Glu Cys Thr Pro Asn Ile			
250	255	260	
GAT GGA CCA AAT GCT AAA TCT GTT CAG AGA GAG CAA AGC TTA CAC TCC			929
Asp Gly Pro Asn Ala Lys Ser Val Gln Arg Glu Gln Ser Leu His Ser			
265	270	275	280
TTT CAT ACG CTT TTC TGT AGG CGA TGT TTT AAA TAT GAC TGC TTC CTA			977
Phe His Thr Leu Phe Cys Arg Arg Cys Phe Lys Tyr Asp Cys Phe Leu			
285	290	295	
CAT CCT TTT CAT GCA ACA CCC AAC ACT TAT AAG CGG AAG AAC ACA GAA			1025
His Pro Phe His Ala Thr Pro Asn Thr Tyr Lys Arg Lys Asn Thr Glu			
300	305	310	
ACA GCT CTA GAC AAC AAA CCT TGT GGA CCA CAG TGT TAC CAG CAT TTG			1073
Thr Ala Leu Asp Asn Lys Pro Cys Gly Pro Gln Cys Tyr Gln His Leu			
315	320	325	
GAG GGA GCA AAG GAG TTT GCT GCT GCT CTC ACC GCT GAG CGG ATA AAG			1121
Glu Gly Ala Lys Glu Phe Ala Ala Ala Leu Thr Ala Glu Arg Ile Lys			
330	335	340	
ACC CCA CCA AAA CGT CCA GGA GGC CGC AGA AGA GGA CGG CTT CCC AAT			1169
Thr Pro Pro Lys Arg Pro Gly Gly Arg Arg Arg Gly Arg Leu Pro Asn			
345	350	355	360

AAC AGT AGC AGG CCC AGC ACC CCC ACC ATT AAT GTG CTG GAA TCA AAG 1217
 Asn Ser Ser Arg Pro Ser Thr Pro Thr Ile Asn Val Leu Glu Ser Lys

365 370 375

GAT ACA GAC AGT GAT AGG GAA GCA GGG ACT GAA ACC GGG GCA GAG AAC 1265
 Asp Thr Asp Ser Asp Arg Glu Ala Gly Thr Glu Thr Gly Gly Glu Asn

380 385 390

AAT GAT AAA GAA GAA GAA GAG AAG AAA GAT GAA ACT TCG AGC TCC TCT 1313
 Asn Asp Lys Glu Glu Glu Glu Lys Lys Asp Glu Thr Ser Ser Ser Ser

395 400 405

GAA GCA AAT TCT CGG TGT CAA ACA CCA ATA AAG ATG AAG CCA AAT ATT 1361
 Glu Ala Asn Ser Arg Cys Gln Thr Pro Ile Lys Met Lys Pro Asn Ile

410 415 420

GAA CCT CCT GAG AAT GTG GAG TGG AGT GGT GCT GAA GCC TCA ATG TTT 1409
 Glu Pro Pro Glu Asn Val Glu Trp Ser Gly Ala Glu Ala Ser Met Phe

425 430 435 440

AGA GTC CTC ATT GGC ACT TAC TAT GAC AAT TTC TGT GCC ATT GCT AGG 1457
 Arg Val Leu Ile Gly Thr Tyr Tyr Asp Asn Phe Cys Ala Ile Ala Arg

445 450 455

TTA ATT GGG ACC AAA ACA TGT AGA CAG GTG TAT GAG TTT AGA GTC AAA 1505
 Leu Ile Gly Thr Lys Thr Cys Arg Gln Val Tyr Glu Phe Arg Val Lys

460 465 470

GAA TCT AGC ATC ATA GCT CCA GCT CCC GCT GAG GAT GTG GAT ACT CCT 1553
 Glu Ser Ser Ile Ile Ala Pro Ala Pro Ala Glu Asp Val Asp Thr Pro

475 480 485

CCA AGG AAA AAG AAG AGG AAA CAC CGG TTG TGG GCT GCA CAC TGC AGA 1601
 Pro Arg Lys Lys Lys Arg Lys His Arg Leu Trp Ala Ala His Cys Arg

490 495 500

AAG ATA CAG CTG AAA AAG GAC GGC TCC TCT AAC CAT GTT TAC AAC TAT 1649
 Lys Ile Gln Leu Lys Lys Asp Gly Ser Ser Asn His Val Tyr Asn Tyr

505 510 515 520

CAA CCC TGT GAT CAT CCA CGG CAG CCT TGT GAC AGT TCG TGC CCT TGT 1697
 Gln Pro Cys Asp His Pro Arg Gln Pro Cys Asp Ser Ser Cys Pro Cys

525 530 535

GTG ATA GCA CAA AAT TTT TGT GAA AAG TTT TGT CAA TGT AGT TCA GAG 1745
 Val Ile Ala Gln Asn Phe Cys Glu Lys Phe Cys Gln Cys Ser Ser Glu

540 545 550

TGT CAA AAC CGC TTT CCG GGA TGC CGC TGC AAA GCA CAG TGC AAC ACC 1793
 Cys Gln Asn Arg Phe Pro Gly Cys Arg Cys Lys Ala Gln Cys Asn Thr

555 560 565

AAG CAG TGC CCG TGC TAC CTG GCT GTC CGA GAG TGT GAC CCT GAC CTC 1841
 Lys Gln Cys Pro Cys Tyr Leu Ala Val Arg Glu Cys Asp Pro Asp Leu

570 575 580

TGT CTT ACT TGT GGA GCC GCT GAC CAT TGG GAC AGT AAA AAT GTG TCC 1889
 Cys Leu Thr Cys Gly Ala Ala Asp His Trp Asp Ser Lys Asn Val Ser
 585 590 595 600

TGC AAG AAC TGC AGT ATT CAG CGG GGC TCC AAA AAG CAT CTA TTG CTG 1937
 Cys Lys Asn Cys Ser Ile Gln Arg Gly Ser Lys Lys His Leu Leu Leu
 605 610 615

GCA CCA TCT GAC GTG GCA GGC TGG GGG ATT TTT ATC AAA GAT CCT GTG 1985
 Ala Pro Ser Asp Val Ala Gly Trp Gly Ile Phe Ile Lys Asp Pro Val
 620 625 630

CAG AAA AAT GAA TTC ATC TCA GAA TAC TGT GGA GAG ATT ATT TCT CAA 2033
 Gln Lys Asn Glu Phe Ile Ser Glu Tyr Cys Gly Glu Ile Ile Ser Gln
 635 640 645

GAT GAA GCT GAC AGA AGA GGG AAA GTG TAT GAT AAA TAC ATG TGC AGC 2081
 Asp Glu Ala Asp Arg Arg Gly Lys Val Tyr Asp Lys Tyr Met Cys Ser
 650 655 660

TTT CTG TTC AAC TTG AAC AAT GAT TTT GTG GTG GAT GCA ACC CGC AAG 2129
 Phe Leu Phe Asn Leu Asn Asn Asp Phe Val Val Asp Ala Thr Arg Lys
 665 670 675 680

GGT AAC AAA ATT CGT TTT GCA AAT CAT TCG GTA AAT CCA AAC TGC TAT 2177
 Gly Asn Lys Ile Arg Phe Ala Asn His Ser Val Asn Pro Asn Cys Tyr
 685 690 695

GCA AAA GTT ATG ATG GTT AAC GGT GAT CAC AGG ATA GGT ATT TTT GCC 2225

Ala Lys Val Met Met Val Asn Gly Asp His Arg Ile Gly Ile Phe Ala
 700 705 710

AAG AGA GCC ATC CAG ACT GGC GAA GAG CTG TTT TTT GAT TAC AGA TAC 2273
 Lys Arg Ala Ile Gln Thr Gly Glu Glu Leu Phe Phe Asp Tyr Arg Tyr
 715 720 725

AGC CAG GCT GAT GCC CTG AAG TAT GTC GGC ATC GAA AGA GAA ATG GAA 2321
 Ser Gln Ala Asp Ala Leu Lys Tyr Val Gly Ile Glu Arg Glu Met Glu
 730 735 740

ATC CCT TGA CATCTGCTAC CTCCTCCCC TCCTCTGAAA CAGCTGCCTT 2370
 Ile Pro *
 745

AGCTTCAGGA ACCTCGAGTA CTGTGGGCAA TTTAGAAAAA GAACATGCAG TTTGAAATTC 2430

TGAATTTGCA AAGTACTGTA AGAATAATTT ATAGTAATGA GTTTAAAAAT CAACTTTTTA 2490

TTGCCTTCTC ACCAGCTGCA AAGTGTTTTG TACCAGTGAA TTTTTCGAAT AATGCAGTAT 2550

GGTACATTTT TCAACTTTGA ATAAAGAATA CTTGAACTTG TCAAAAAAAA 2600

(2) 配列番号：2 の情報：

(i) 配列の特徴：

(A) 長さ：747 アミノ酸

(B) 型：アミノ酸

(D) トポロジー：直鎖状

(ii)分子の種類：タンパク質

(xi)配列：配列番号：2：

Met Gly Gln Thr Gly Lys Lys Ser Glu Lys Gly Pro Val Cys Trp Arg
1 5 10 15

Lys Arg Val Lys Ser Glu Tyr Met Arg Leu Arg Gln Leu Lys Arg Phe
20 25 30

Arg Arg Ala Asp Glu Val Lys Ser Met Phe Ser Ser Asn Arg Gln Lys
35 40 45

Ile Leu Glu Arg Thr Glu Ile Leu Asn Gln Glu Trp Lys Gln Arg Arg
50 55 60

Ile Gln Pro Val His Ile Leu Thr Ser Val Ser Ser Leu Arg Gly Thr
65 70 75 80

Arg Glu Cys Ser Val Thr Ser Asp Leu Asp Phe Pro Thr Gln Val Ile
85 90 95

Pro Leu Lys Thr Leu Asn Ala Val Ala Ser Val Pro Ile Met Tyr Ser
100 105 110

Trp Ser Pro Leu Gln Gln Asn Phe Met Val Glu Asp Glu Thr Val Leu
115 120 125

His Asn Ile Pro Tyr Met Gly Asp Glu Val Leu Asp Gln Asp Gly Thr
130 135 140

Phe Ile Glu Glu Leu Ile Lys Asn Tyr Asp Gly Lys Val His Gly Asp
145 150 155 160

Arg Glu Cys Gly Phe Ile Asn Asp Glu Ile Phe Val Glu Leu Val Asn
165 170 175

Ala Leu Gly Gln Tyr Asn Asp Asp Asp Asp Asp Asp Gly Asp Asp
180 185 190

Pro Glu Glu Arg Glu Glu Lys Gln Lys Asp Leu Glu Asp His Arg Asp
195 200 205

Asp Lys Glu Ser Arg Pro Pro Arg Lys Phe Pro Ser Asp Lys Ile Phe
210 215 220

Glu Ala Ile Ser Ser Met Phe Pro Asp Lys Gly Thr Ala Glu Glu Leu
225 230 235 240

Lys Glu Lys Tyr Lys Glu Leu Thr Glu Gln Gln Leu Pro Gly Ala Leu
245 250 255

Pro Pro Glu Cys Thr Pro Asn Ile Asp Gly Pro Asn Ala Lys Ser Val
260 265 270

Gln Arg Glu Gln Ser Leu His Ser Phe His Thr Leu Phe Cys Arg Arg
275 280 285

Cys Phe Lys Tyr Asp Cys Phe Leu His Pro Phe His Ala Thr Pro Asn

290	295	300	
Thr Tyr Lys Arg Lys Asn Thr Glu Thr Ala Leu Asp Asn Lys Pro Cys			
305	310	315	320
Gly Pro Gln Cys Tyr Gln His Leu Glu Gly Ala Lys Glu Phe Ala Ala			
	325	330	335
Ala Leu Thr Ala Glu Arg Ile Lys Thr Pro Pro Lys Arg Pro Gly Gly			
	340	345	350
Arg Arg Arg Gly Arg Leu Pro Asn Asn Ser Ser Arg Pro Ser Thr Pro			
	355	360	365
Thr Ile Asn Val Leu Glu Ser Lys Asp Thr Asp Ser Asp Arg Glu Ala			
	370	375	380
Gly Thr Glu Thr Gly Gly Glu Asn Asn Asp Lys Glu Glu Glu Glu Lys			
385	390	395	400
Lys Asp Glu Thr Ser Ser Ser Ser Glu Ala Asn Ser Arg Cys Gln Thr			
	405	410	415
Pro Ile Lys Met Lys Pro Asn Ile Glu Pro Pro Glu Asn Val Glu Trp			
	420	425	430
Ser Gly Ala Glu Ala Ser Met Phe Arg Val Leu Ile Gly Thr Tyr Tyr			
	435	440	445

Asp Asn Phe Cys Ala Ile Ala Arg Leu Ile Gly Thr Lys Thr Cys Arg

450

455

460

Gln Val Tyr Glu Phe Arg Val Lys Glu Ser Ser Ile Ile Ala Pro Ala

465

470

475

480

Pro Ala Glu Asp Val Asp Thr Pro Pro Arg Lys Lys Lys Arg Lys His

485

490

495

Arg Leu Trp Ala Ala His Cys Arg Lys Ile Gln Leu Lys Lys Asp Gly

500

505

510

Ser Ser Asn His Val Tyr Asn Tyr Gln Pro Cys Asp His Pro Arg Gln

515

520

525

Pro Cys Asp Ser Ser Cys Pro Cys Val Ile Ala Gln Asn Phe Cys Glu

530

535

540

Lys Phe Cys Gln Cys Ser Ser Glu Cys Gln Asn Arg Phe Pro Gly Cys

545

550

555

560

Arg Cys Lys Ala Gln Cys Asn Thr Lys Gln Cys Pro Cys Tyr Leu Ala

565

570

575

Val Arg Glu Cys Asp Pro Asp Leu Cys Leu Thr Cys Gly Ala Ala Asp

580

585

590

His Trp Asp Ser Lys Asn Val Ser Cys Lys Asn Cys Ser Ile Gln Arg

595

600

605

Gly Ser Lys Lys His Leu Leu Leu Ala Pro Ser Asp Val Ala Gly Trp
610 615 620

Gly Ile Phe Ile Lys Asp Pro Val Gln Lys Asn Glu Phe Ile Ser Glu
625 630 635 640

Tyr Cys Gly Glu Ile Ile Ser Gln Asp Glu Ala Asp Arg Arg Gly Lys
645 650 655

Val Tyr Asp Lys Tyr Met Cys Ser Phe Leu Phe Asn Leu Asn Asn Asp
660 665 670

Phe Val Val Asp Ala Thr Arg Lys Gly Asn Lys Ile Arg Phe Ala Asn
675 680 685

His Ser Val Asn Pro Asn Cys Tyr Ala Lys Val Met Met Val Asn Gly
690 695 700

Asp His Arg Ile Gly Ile Phe Ala Lys Arg Ala Ile Gln Thr Gly Glu
705 710 715 720

Glu Leu Phe Phe Asp Tyr Arg Tyr Ser Gln Ala Asp Ala Leu Lys Tyr
725 730 735

Val Gly Ile Glu Arg Glu Met Glu Ile Pro *
740 745

(2) 配列番号：3 の情報：

(i) 配列の特徴：

(A) 長さ：2732 塩基対

(B) 型：ヌクレオチド

(C) 鎖：一本鎖

(D) トポロジー：直鎖状

(ii) 分子の種類：cDNA to mRNA

(iii) ハイボセティカル：No

(iv) アンチセンス：No

(vi) 起源：

(A) 生物名：ヒト

(G) 細胞の種類：B-cell

(ix) 特徴：

(A) 名称／キー：5' UTR

(B) 存在位置：1..44

(ix) 特徴：

(A) 名称／キー：CDS

(B) 存在位置：45..1283

(ix) 特徴：

(A) 名称／キー：3' UTR

(B) 存在位置：1284..2732

(xi) 配列：配列番号：3：

TCGCGAGGCC GGCTAGGCCC GAATGTCGTT AGCCGTGGGG AAAG ATG GCG GAA AAT 56
Met Ala Glu Asn
750

TTA AAA GGC TGC AGC GTG TGT TGC AAG TCT TCT TGG AAT CAG CTG CAG 104
Leu Lys Gly Cys Ser Val Cys Cys Lys Ser Ser Trp Asn Gln Leu Gln

755	760	765	
GAC CTG TGC CGC CTG GCC AAG CTC TCC TGC CCT GCC CTC GGT ATC TCT			152
Asp Leu Cys Arg Leu Ala Lys Leu Ser Cys Pro Ala Leu Gly Ile Ser			
770	775	780	
AAG AGG AAC CTC TAT GAC TTT GAA GTC GAG TAC CTG TGC GAT TAC AAG			200
Lys Arg Asn Leu Tyr Asp Phe Glu Val Glu Tyr Leu Cys Asp Tyr Lys			
785	790	795	
AAG ATC CGC GAA CAG GAA TAT TAC CTG GTG AAA TGG CGT GGA TAT CCA			248
Lys Ile Arg Glu Gln Glu Tyr Tyr Leu Val Lys Trp Arg Gly Tyr Pro			
800	805	810	815
GAC TCA GAG AGC ACC TGG GAG CCA CGG CAG AAT CTC AAG TGT GTG CGT			296
Asp Ser Glu Ser Thr Trp Glu Pro Arg Gln Asn Leu Lys Cys Val Arg			
820	825	830	
ATC CTC AAG CAG TTC CAC AAG GAC TTA GAA AGG GAG CTG CTC CGG CGG			344
Ile Leu Lys Gln Phe His Lys Asp Leu Glu Arg Glu Leu Leu Arg Arg			
835	840	845	
CAC CAC CGG TCA AAG ACC CCC CGG CAC CTG GAC CCA AGC TTG GCC AAC			392
His His Arg Ser Lys Thr Pro Arg His Leu Asp Pro Ser Leu Ala Asn			
850	855	860	
TAC CTG GTG CAG AAG GCC AAG CAG AGG CGG GCG CTC CGT CGC TGG GAG			440
Tyr Leu Val Gln Lys Ala Lys Gln Arg Arg Ala Leu Arg Arg Trp Glu			
865	870	875	

CAG GAG CTC AAT GCC AAG CGC AGC CAT CTG GGA CGC ATC ACT GTA GAG 488
 Gln Glu Leu Asn Ala Lys Arg Ser His Leu Gly Arg Ile Thr Val Glu
 880 885 890 895

AAT GAG GTG GAC CTG GAC GGC CCT CCG CGG GCC TTC GTG TAC ATC AAT 536
 Asn Glu Val Asp Leu Asp Gly Pro Pro Arg Ala Phe Val Tyr Ile Asn
 900 905 910

GAG TAC CGT GTT GGT GAG GGC ATC ACC CTC AAC CAG GTG GCT GTG GGC 584
 Glu Tyr Arg Val Gly Glu Gly Ile Thr Leu Asn Gln Val Ala Val Gly
 915 920 925

TGC GAG TGC CAG GAC TGT CTG TGG GCA CCC ACT GGA GGC TGC TGC CCG 632
 Cys Glu Cys Gln Asp Cys Leu Trp Ala Pro Thr Gly Gly Cys Cys Pro
 930 935 940

GGG GCG TCA CTG CAC AAG TTT GCC TAC AAT GAC CAG GGC CAG GTG CGG 680
 Gly Ala Ser Leu His Lys Phe Ala Tyr Asn Asp Gln Gly Gln Val Arg
 945 950 955

CTT CGA GCC GGG CTG CCC ATC TAC GAG TGC AAC TCC CGC TGC CGC TGC 728
 Leu Arg Ala Gly Leu Pro Ile Tyr Glu Cys Asn Ser Arg Cys Arg Cys
 960 965 970 975

GGC TAT GAC TGC CCA AAT CGT GTG GTA CAG AAG GGT ATC CGA TAT GAC 776
 Gly Tyr Asp Cys Pro Asn Arg Val Val Gln Lys Gly Ile Arg Tyr Asp
 980 985 990

CTC TGC ATC TTC CGG ACG GAT GAT GGG CGT GGC TGG GGC GTC CGC ACC	824
Leu Cys Ile Phe Arg Thr Asp Asp Gly Arg Gly Trp Gly Val Arg Thr	
995 1000 1005	
CTG GAG AAG ATT CGC AAG AAC AGC TTC GTC ATG GAG TAC GTG GGA GAG	872
Leu Glu Lys Ile Arg Lys Asn Ser Phe Val Met Glu Tyr Val Gly Glu	
1010 1015 1020	
ATC ATT ACC TCA GAG GAG GCA GAG CGG CGG GGC CAG ATC TAC GAC CGT	920
Ile Ile Thr Ser Glu Glu Ala Glu Arg Arg Gly Gln Ile Tyr Asp Arg	
1025 1030 1035	
CAG GGC GCC ACC TAC CTC TTT GAC CTG GAC TAC GTG GAG GAC CTG TAC	968
Gln Gly Ala Thr Tyr Leu Phe Asp Leu Asp Tyr Val Glu Asp Val Tyr	
1040 1045 1050 1055	
ACC GTG GAT GCC GCC TAC TAT GGC AAC ATC TCC CAC TTT GTC AAC CAC	1016
Thr Val Asp Ala Ala Tyr Tyr Gly Asn Ile Ser His Phe Val Asn His	
1060 1065 1070	
AGT TGT GAC CCC AAC CTG CAG GTG TAC AAC GTC TTC ATA GAC AAC CTT	1064
Ser Cys Asp Pro Asn Leu Gln Val Tyr Asn Val Phe Ile Asp Asn Leu	
1075 1080 1085	
GAC GAG CGG CTG CCC CGC ATC GCT TTC TTT GCC ACA AGA ACC ATC CGG	1112
Asp Glu Arg Leu Pro Arg Ile Ala Phe Phe Ala Thr Arg Thr Ile Arg	
1090 1095 1100	

GCA GGC GAG GAG CTC ACC TTT GAT TAC AAC ATG CAA GTG GAC CCC GTG	1160
Ala Gly Glu Glu Leu Thr Phe Asp Tyr Asn Met Gln Val Asp Pro Val	
1105 1110 1115	
GAC ATG GAG AGC ACC CGC ATG GAC TCC AAC TTT GGC CTG GCT GGG CTC	1208
Asp Met Glu Ser Thr Arg Met Asp Ser Asn Phe Gly Leu Ala Gly Leu	
1120 1125 1130 1135	
CCT GGC TCC CCT AAG AAG CGG GTC CGT ATT GAA TGC AAG TGT GGG ACT	1256
Pro Gly Ser Pro Lys Lys Arg Val Arg Ile Glu Cys Lys Cys Gly Thr	
1140 1145 1150	
GAG TCC TGC CGC AAA TAC CTC TTC TAG CCCTTAGAAG TCTGAGGCCA	1303
Glu Ser Cys Arg Lys Tyr Leu Phe *	
1155 1160	
GACTGACTGA GGGGGCCTGA AGCTACATGC ACCTCCCCCA CTGCTGCCCT CCTGTGACA	1363
ATGACTGCCA GGGCCTCGCC TGCCTCCACC TGCCCCCACC TGCTCCTACC TGCTCTACCT	1423
TCAGGGCTGT GGGCGTGGTG AGGACCGACT CCAGGAGTCC CCTTTCCTG TCCCAGCCCC	1483
ATCTGTGGGT TGCACTTACA AACCCCCACC CACCTTCAGA AATAGTTTTT CAACATCAAG	1543
ACTCTCTGTC GTTGGGATTC ATGGCCTATT AAGGAGGTCC AAGGGGTGAG TCCCAACCCA	1603
GGCCAGAAT ATATTTGTTT TTGCACCTGC TTCTGCCTGG AGATTGAGGG GTCTGCTGCA	1663
GGCCTCCTCC CTGCTGCCCC AAAGGTATGG GGAAGCAACC CCAGAGCAGG CAGACATCAG	1723

AGGCCAGAGT GCCTAGCCCG ACATGAAGCT GGTTCCCCAA CCACAGAAAC TTTGTACTAG	1783
TGAAAGAAAG GGGTCCCTGG CCTACGGGCT GAGGCTGGTT TCTGCTCGTG CTTACAGTGC	1843
TGGGTAGTGT TGGCCCTAAG AGCTGTAGGG TCTCTTCTTC AGGGCTGCAT ATCTGAGAAG	1903
TGGATGCCCC CATGCCACTG GAAGGGAAGT GGGTGTCCAT GGGCCACTGA GCAGTGAGAG	1963
GAAGGCAGTG CAGAGCTGGC CAGCCCTGGA GGTAGGCTGG GACCAAGCTC TGCCTTCACA	2023
GTGCAGTGAA GGTACCTAGG GCTCTTGGGA GCTCTGCGGT TGCTAGGGGC CCTGACCTGG	2083
GGTGTCATGA CCGCTGACAC CACTCAGAGC TGGAACCAAG ATCTAGATAG TCCCTAGATA	2143
GCACTTAGGA CAAGAATGTG CATTGATGGG GTGGTGATGA GGTGCCAGGC ACTAGGTAGA	2203
GCACCTGGTC CACGTGGATT GTCTCAGGGA AGCCTTGAAA ACCACGGAGG TGGATGCCAG	2263
GAAAGGGCCC ATGTGCCAGA AGGCCAACTA CAGGCCAAGA ATTGGGGGTG GGGGAGATGG	2323
CTTCCCCACT ATGGGATGAC GAGGCGAGAG GGAAGCCCTT GCTGCCTGCC ATTCCCAGAC	2383
CCCAGCCCTT TGTGCTCACC CTGGTTCCAC TGGTCTCAA AGTCACCTGC CTACAAATGT	2443
ACAAAAGGCG AAGGTTCTGA TGGCTGCCTT GCTCCTTGCT CCCCCACCCC CTGTGAGGAC	2503
TTCTCTAGGA AGTCCTTCCT GACTACCTGT GCCCAGAGTG CCCCTACATG AACTGTATG	2563

CCCTGCTATC AGATGCCAGA TCTATGTGTC TGTCTGTGTG TCCATCCCGC CGGCCCCCCA 2623

GACTAACCTC CAGGCATGGA CTGAATCTGG TTCTCCTCTT GTACACCCCT CAACCCATG 2683

CAGCCTGGAG TGGGCATCAA TAAATGAAC TGTCGACTGA AAAAAAAAAA 2732

(2) 配列番号 : 4 の情報 :

(i) 配列の特徴 :

(A) 長さ : 413 アミノ酸

(B) 型 : アミノ酸

(D) トポロジー : 直鎖状

(ii) 分子の種類 : タンパク質

(xi) 配列 : 配列番号 : 4 :

Met Ala Glu Asn Leu Lys Gly Cys Ser Val Cys Cys Lys Ser Ser Trp

1 5 10 15

Asn Gln Leu Gln Asp Leu Cys Arg Leu Ala Lys Leu Ser Cys Pro Ala

20 25 30

Leu Gly Ile Ser Lys Arg Asn Leu Tyr Asp Phe Glu Val Glu Tyr Leu

35 40 45

Cys Asp Tyr Lys Lys Ile Arg Glu Gln Glu Tyr Tyr Leu Val Lys Trp

50 55 60

Arg Gly Tyr Pro Asp Ser Glu Ser Thr Trp Glu Pro Arg Gln Asn Leu

65 70 75 80

Lys Cys Val Arg Ile Leu Lys Gln Phe His Lys Asp Leu Glu Arg Glu

85

90

95

Leu Leu Arg Arg His His Arg Ser Lys Thr Pro Arg His Leu Asp Pro

100

105

110

Ser Leu Ala Asn Tyr Leu Val Gln Lys Ala Lys Gln Arg Arg Ala Leu

115

120

125

Arg Arg Trp Glu Gln Glu Leu Asn Ala Lys Arg Ser His Leu Gly Arg

130

135

140

Ile Thr Val Glu Asn Glu Val Asp Leu Asp Gly Pro Pro Arg Ala Phe

145

150

155

160

Val Tyr Ile Asn Glu Tyr Arg Val Gly Glu Gly Ile Thr Leu Asn Gln

165

170

175

Val Ala Val Gly Cys Glu Cys Gln Asp Cys Leu Trp Ala Pro Thr Gly

180

185

190

Gly Cys Cys Pro Gly Ala Ser Leu His Lys Phe Ala Tyr Asn Asp Gln

195

200

205

Gly Gln Val Arg Leu Arg Ala Gly Leu Pro Ile Tyr Glu Cys Asn Ser

210

215

220

Arg Cys Arg Cys Gly Tyr Asp Cys Pro Asn Arg Val Val Gln Lys Gly

225

230

235

240

Ile Arg Tyr Asp Leu Cys Ile Phe Arg Thr Asp Asp Gly Arg Gly Trp
245 250 255

Gly Val Arg Thr Leu Glu Lys Ile Arg Lys Asn Ser Phe Val Met Glu
260 265 270

Tyr Val Gly Glu Ile Ile Thr Ser Glu Glu Ala Glu Arg Arg Gly Gln
275 280 285

Ile Tyr Asp Arg Gln Gly Ala Thr Tyr Leu Phe Asp Leu Asp Tyr Val
290 295 300

Glu Asp Val Tyr Thr Val Asp Ala Ala Tyr Tyr Gly Asn Ile Ser His
305 310 315 320

Phe Val Asn His Ser Cys Asp Pro Asn Leu Gln Val Tyr Asn Val Phe
325 330 335

Ile Asp Asn Leu Asp Glu Arg Leu Pro Arg Ile Ala Phe Phe Ala Thr
340 345 350

Arg Thr Ile Arg Ala Gly Glu Glu Leu Thr Phe Asp Tyr Asn Met Gln
355 360 365

Val Asp Pro Val Asp Met Glu Ser Thr Arg Met Asp Ser Asn Phe Gly
370 375 380

Leu Ala Gly Leu Pro Gly Ser Pro Lys Lys Arg Val Arg Ile Glu Cys

385

390

395

400

Lys Cys Gly Thr Glu Ser Cys Arg Lys Tyr Leu Phe *

405

410

(2) 配列番号 : 5 の情報 :

(i) 配列の特徴 :

(A) 長さ : 489 塩基対

(B) 型 : ヌクレオチド

(C) 鎖 : 一本鎖

(D) トポロジー : 直鎖状

(ii) 分子の種類 : cDNA to mRNA

(iii) ハイボセティカル : No

(iv) アンチセンス : No

(vi) 起源 :

(A) 生物名 : ヒト

(G) 細胞の種類 : B-cell

(ix) 特徴 :

(A) 名称 / キー : CDS

(B) 存在位置 : 1..341

(C) 他の特徴 : 部分配列、配列番号 1 の配列に相同

(ix) 特徴 :

(A) 名称 / キー : ハイボセティカル非コード領域

(B) 存在位置 : 342..489

(xi) 配列 : 配列番号 : 5 :

A CTC ACC TGT GGG GCC TCA GAG CAC TGG GAC TGC AAG GTG GTT TCC

46

Leu Thr Cys Gly Ala Ser Glu His Trp Asp Cys Lys Val Val Ser

1

5

10

15

TGT AAA AAC TGC AGC ATC CAG CGT GGA CTT AAG AAG CAC CTG CTG CTG 94
 Cys Lys Asn Cys Ser Ile Gln Arg Gly Leu Lys Lys His Leu Leu Leu

20 25 30

GCC CCC TCT GAT GTG GCC GGA TGG GGC ACC TTC ATA AAG GAG TCT GTG 142
 Ala Pro Ser Asp Val Ala Gly Trp Gly Thr Phe Ile Lys Glu Ser Val

35 40 45

CAG AAG AAC GAA TTC ATT TCT GAA TAC TGT GGT GAG CTC ATC TCT CAG 190
 Gln Lys Asn Glu Phe Ile Ser Glu Tyr Cys Gly Glu Leu Ile Ser Gln

50 55 60

GAT GAG GCT GAT CGA CGC GGA AAG GTC TAT GAC AAA TAC ATG TCC AGC 238
 Asp Glu Ala Asp Arg Arg Gly Lys Val Tyr Asp Lys Tyr Met Ser Ser

65 70 75

TTC CTC TTC AAC CTC AAT AAT GAT TTT GTA GTG GAT GCT ACT CGG AAA 286
 Phe Leu Phe Asn Leu Asn Asn Asp Phe Val Val Asp Ala Thr Arg Lys

80 85 90 95

GGA AAC AAA ATT CGA TTT GCA AAT CAT TCA GTG AAT CCC AAC TGT TAT 334
 Gly Asn Lys Ile Arg Phe Ala Asn His Ser Val Asn Pro Asn Cys Tyr

100 105 110

GCC AAA G GTGAGTCCCA GTAACCTGGG AGGTGGGGTG GCGGATGGAT GCCTCTTTAC 391
 Ala Lys

TGTGATTTC ATTGTTTGT GAACATTTTC CTTAGCTGAG CTATCTTTTG 441

TCCAAAGATA ATCATGATTA ATATCTGGTA TCATTTTAGG CCCCTCTC.

489

(2) 配列番号 : 6 の情報 :

(i) 配列の特徴 :

(A) 長さ : 113 アミノ酸

(B) 型 : アミノ酸

(D) トポロジー : 直鎖状

(E) WEITERES MERKMAL: Teilsequenz, Homologie zu SEQ ID NO. 2

(ii) 分子の種類 : タンパク質

(xi) 配列 : 配列番号 : 6 :

Leu Thr Cys Gly Ala Ser Glu His Trp Asp Cys Lys Val Val Ser Cys
1 5 10 15

Lys Asn Cys Ser Ile Gln Arg Gly Leu Lys Lys His Leu Leu Leu Ala
20 25 30

Pro Ser Asp Val Ala Gly Trp Gly Thr Phe Ile Lys Glu Ser Val Gln
35 40 45

Lys Asn Glu Phe Ile Ser Glu Tyr Cys Gly Glu Leu Ile Ser Gln Asp
50 55 60

Glu Ala Asp Arg Arg Gly Lys Val Tyr Asp Lys Tyr Met Ser Ser Phe
65 70 75 80

Leu Phe Asn Leu Asn Asn Asp Phe Val Val Asp Ala Thr Arg Lys Gly
85 90 95

Asn Lys Ile Arg Phe Ala Asn His Ser Val Asn Pro Asn Cys Tyr Ala

100

105

110

Lys

(2) 配列番号：7 の情報：

(i) 配列の特徴：

- (A) 長さ：20 塩基対
- (B) 型：ヌクレオチド
- (C) 鎖：一本鎖
- (D) トポロジー：直鎖状

(ii) 分子の種類：DNA

(iii) ハイポセティカル：No

(iv) アンチセンス：No

(ix) 特徴：

合成アダプター分子

(xi) 配列：配列番号：7：

AATTCTCGAG CTCGTCGACA

20

(2) 配列番号：8 の情報：

(i) 配列の特徴：

- (A) 長さ：31 塩基対
- (B) 型：ヌクレオチド
- (C) 鎖：一本鎖
- (D) トポロジー：直鎖状

(ii) 分子の種類：DNA

(iii) ハイポセティカル：No

(iv)アンチセンス : No

(ix)特徴 :

合成プライマー分子

(xi)配列 : 配列番号 : 8 :

ACTGAATTCG GCTGGGGCAT CTTTCTTAAG G

31

(2) 配列番号 : 9 の情報 :

(i)配列の特徴 :

(A) 長さ : 31 塩基対

(B) 型 : ヌクレオチド

(C) 鎖 : 一本鎖

(D) トポロジー : 直鎖状

(ii)分子の種類 : DNA

(iii)ハイボセティカル : No

(iv)アンチセンス : No

(ix)特徴 :

合成プライマー分子

(xi)配列 : 配列番号 : 9 :

ACTCTAGACA ATTTCCATTT CACGCTCTAT G

31

(2) 配列番号 : 10 の情報 :

(i)配列の特徴 :

(A) 長さ : 30塩基対

(B) 型 : ヌクレオチド

(C) 鎖 : 一本鎖

(D) トポロジー : 直鎖状

(ii)分子の種類 : DNA

(iii)ハイボセティカル: No

(iv)アンチセンス: No

(ix)特徴:

合成プライマー分子

(xi)配列: 配列番号: 10:

ATATAGTACT TCAAGTCCAT TCAAAAGAGG

30

(2) 配列番号: 11の情報:

(i)配列の特徴:

(A) 長さ: 29 塩基対

(B) 型: ヌクレオチド

(C) 鎖: 一本鎖

(D) トポロジー: 直鎖状

(ii)分子の種類: DNA

(iii)ハイボセティカル: No

(iv)アンチセンス: No

(ix)特徴:

合成プライマー分子

(xi)配列: 配列番号: 11:

CCAGGTACCG TTGGTGCTGT TTAAGACCG

29

EZH2 —1 MGQTGKKSEKGPVCHRKRVKSEYMRLRQLKRFRRRADEVKSMFSSNRQKIL 50
 E(z) —1MNSTKVPPWKRRVKSEYIKIRQOKRYKRADEIKKAWIRNWDERN 45
 51 ERTEILNQEWKORRIQPVHILTSVSSLRGITRECSVTSDLDFP..TQVIPL 98
 46 HNVQDLYCESKVHQAQPYD....PPHVDCVKRAEVTSYNGIPSGPQKVP 91
 99 KTLNAVASVPIMYSWSPLQONFMVEDEVTLHNIPYMGDEVLDQDGT FIEE 148
 92 CVINAVTPIPTMYTWAPTQONFMVEDEVTLHNIPYMGDEVLDKDGK FIEE 141
 149 LIKNYDGKVGHDRECGFINDEIFVELVNAL..... 178
 142 LIKNYDGKVGHDKDFSMDDAIFVELVHALMRSYSKELEEAAPSTSTAIK 191
 179GQYNDDDDDDGDDPE.....EREKQKQLED.....H 206
 192 TEPLAKSKQGEDDGVVDVADCESPMKLEKTESKGDLDVVEKETEEEPVE 241
 207 RDDKESRPPRK.....FPSDKIFEALSSMFPDKGTAEELKEKYKELTE 249
 242 TEDADVKAPEVEVKDKLPFPAPIIFQAISANFPDKGTAEELKEKYKELTE 291
 250 QQLPGALPPESTPNIDGNKSVQREOSLSFHTLEGRF KYC LHPF 299
 292 HQDPER.PQESTPNIDGIKESVSRRERIMSFHTLEGRF KYC LHL 340
 300 ..HATPNTYKRKNTETALONKEGSPGQYHLEGAKFAAALTAERIKTPP 347
 341 QGHAGPNLOKRRYPKLPFAEFENSLMLIDGMKEKLAA....DSKTPP 386
 348 KRPGCRRRGRLPNNSSRPSTPTINVLESKDTSDREAGTETGENNDKEE 397
 387IDSCNEASSEDNSNSQFSNKDFNA 412
 398 EEKQDET.SSSSEANSRCQTPIKMKPNIEPPENVEWSGAESMFRVLIGT 446
 413 ENSKDNGLTVNSAAVAEINSIMAGMTNITSTOCV.WTGADQALYVLHKV 461
 447 YYDNEIARLIGTKTRQVYEFVRKESIIAPAPAEDVDTPPRKKRKH 496
 462 YLKNIAIAENMLTKTRQVYEFQKEDAEFSFEDLRQDFTPPRKKKKKQ 511
 497 RLMAAHRKIQLKKGSSNHVYNYQEDHPROESSTQVIAQNEK 546
 512 RLNSLRKIQLKKGSSNHVYNYTECHPGHEMCIQTONEK 561
 547 QSSSEQNRFPCKAKQNTKQYLAVEROPDILTGGAADHWS 596
 562 QSSSEQNRFPCKAKQNTKQYLAVEROPDILTGGAADHWS 610
 597 XNVSQNCISIQSGSKHLLAPSDVAITGHEKDEYKONELLSEYGG 646
 611 TKINQVCVORGLKHLLMAPSDIAGCTYKESACQNEFSGYAGEH 660
 647 QDEADRGKVDKVCSELENNDENVDAIKKNTGIRFANHSVAPCA 696
 661 QDEADRGKVDKVCSELENNDENVDAIKKNTGIRFANHSVAPCA 710
 697 KVMVUNGDAIGHAKVALLTGELZEDGKQADAEVYQVGLP 746
 711 KVMVUNGDAIGHAKVALLTGELZEDGKQADAEVYQVGLP 760

Fig. 1

システインに富む

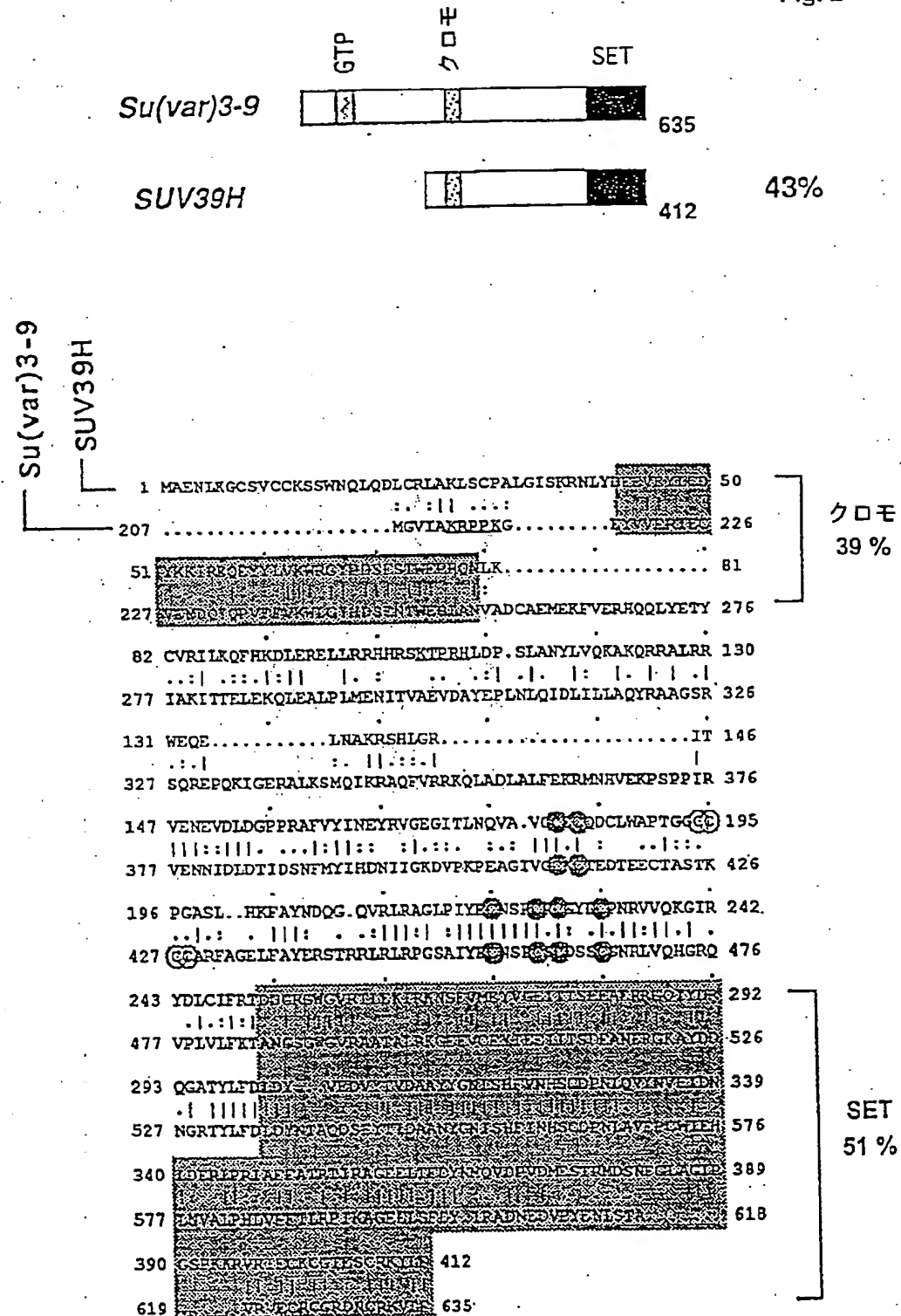
75%

SET

88%

【図2】

Fig. 2



【図 3】

異常な転写産物

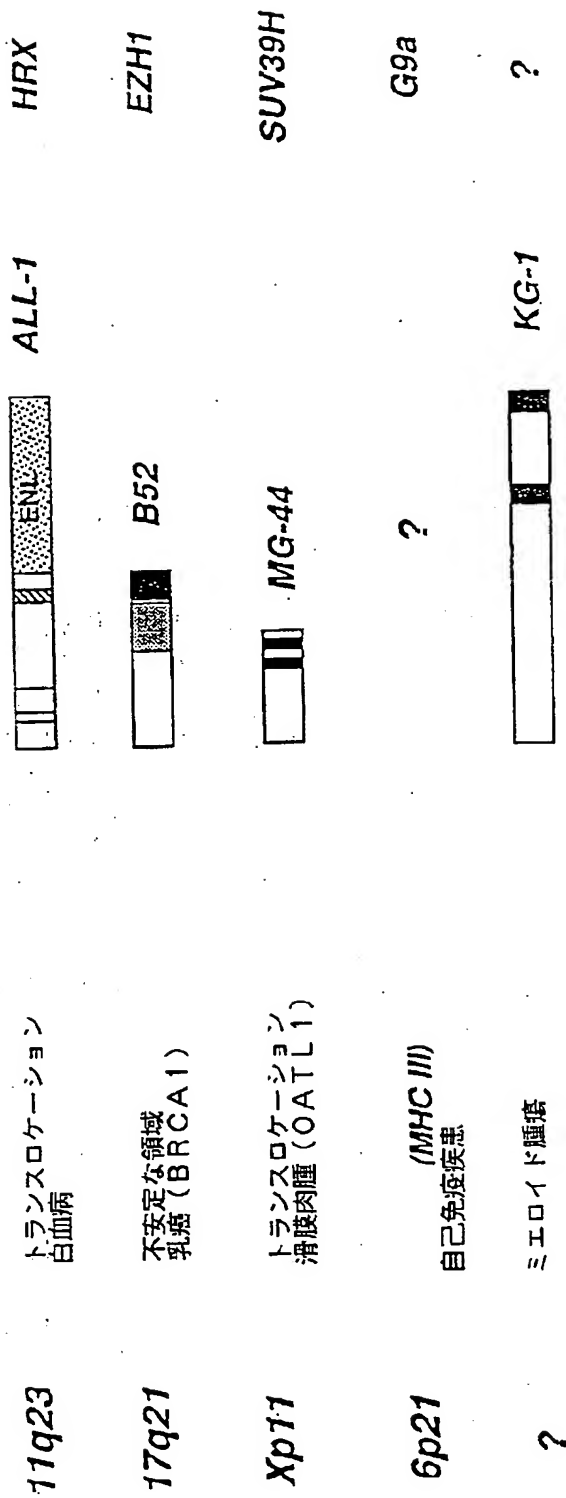
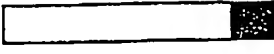



Fig. 3

【図4】

Fig. 4


SETタンパク質ファミリー

S. セレブジアエ YHR119  1080

C. エレガンス C26E6.10  739

シヨウジョウバエ属M.


trx  3751


E(z)  760

Su(var)3-9  635

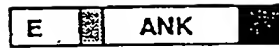
T
E
S

HMG VY
HRX  3969 46%

EZH2  746 61%

SUV39H  412 43%

E
T

G9a  1001
HMG-1

?

[5]

Fig. 5

E(z)	SDIAGWIFL	KEGAQRNEFI	SEYCEIISQ	DEADRGKVY	DK..YMCsFL	50
EZH2	SDVAGWIFI	KDPVQRNEFI	SEYCEIISQ	DEADRGKVY	DK..YMCsFL	
HRX	SPIHGRGLFC	KRNIDAGEMV	IEYAGNVIRS	IOTDKREKY	DSKGIG.CYM	
trx	SHIHGRGLYC	TKDIEAGEMV	IEYAGELIRS	TLTDKRERY	DSRGIG.CYM	
C26	SRIHGWGLYA	MESIAPEDEMI	VEYIGQTIERS	LVAEREKAY	ERRGIGSSYL	
YHR	SAIHNGWGLYA	LDSIAAKEMI	IEYVGERIRQ	PVAEMREKRY	LKNGIGSSYL	
Su3-9	ANGSGWGVRA	ATALRKGEFV	CZYIEELITS	DEANERKAY	DDNG..RTYL	
SUV39H	DDGRGWGVRT	LEKIRKNSFV	MEYVGEIITS	EEAERRGQTY	DRQG..ATYL	
G9a	TAKMGWGVRA	LQTIPOGTFI	CEYVGEIITS	AEAD..V.	.RED..DSYL	
KG-1	TQNKGWGIRC	LDDIAKGSFV	CIYAGKILTD	DFADKEGL.	.EMG..DEYF	

E(z)	ENLN.....	NDFVVDATRK	GNKIREANHS	INPNKYAKVM	MVTGDH....	100
EZH2	ENLN.....	NDFVVDATRK	GNKIREANHS	VNPNKYAKVM	MVNGDH....	
HRX	FRID.....	DSEVVDAIMH	GNRAREINHS	CEPNKYSRVI	NIDGOK....	
trx	FKID.....	DNLVVDATMR	GNAREINHC	CEPNKYSKVV	DILGHK....	
C26	FRID.....	LHHVIDATKR	GNFAREINHS	CEPNKYAKVL	TIEGER....	
YHR	FRVD.....	ENTVIDATKK	GGIAREINEC	CDENCTAKII	KVGRR....	
Su3-9	FDLDYNTAQD	SEYTIIDANY	GNISHEINHS	CDENLAVFPC	WIEHNLVALP	
SUV39H	FDLDY...VE	DVYTVDAAYY	GNISHEINHS	CDENLQVYNY	FIDNLDRLP	
G9a	FDLDNK..DG	EVYCIDARYY	GNISREINHL	CDENIIPVRV	FMLHODLRF	
KG-1	ANLDHI..ES	VEYIIDAKLE	GNLGRYLNHS	CDENLQVYNY	FVDTHDLRF	

△

E(z)	RTGIFAKRAY	QPGHELFEDY	..RYGPTQL K.....	FVGI	EREMEIV*	150
EZH2	RTGIFAKRAY	QPGHELFEDY	..RYSQADAL K.....	YVGI	EREMEIP*	
HRX	HVVFAMRY	YRGHELTIDY	..KFPPIE.DA	SNKLPCNCGA	KKCRKFLN*	
trx	HLIIFAVRY	VOGHELTIDY	..KFPPIE.D.	EKIPCSCGS	KRCRKYLN*	
C26	RVVYSRTTI	KKGHELTIDY	..KFPPIE...	DDKIDCLCGA	KTCRGYLN*	
YHR	REVIALRDI	AASHELTIDY	..KFEREKDD	EERLPCICGA	PNCKGFLN*	
Su3-9	HLVFTLRPI	KAGEELSFEDY	..IRADNEDVP	YENLSTA...	
SUV39H	RTAFFATRI	RAGEELTFEDY	NMQVDPVDME	STRMDSNFG	AGLPGSPKKR	
G9a	RTAFFSSDI	RTGHELGFDY	GDRFW..DIK	SKYFTCQCGS	EKCKHSAEI	
KG-1	WVAFFASKPI	RAGEELTFEDY	NYEVG..SVE	GKELLCCCGA	IECR.....	

E(z)
EZH2

HRX
trx
26
YHR

Su3-9	VRVECRGGRD	NCRKVL*
SUV39H	VRIECKCGTE	SCRKYL*
G9a	ALEQSRLARL	DPHPELLPEL GSLPPVNT*
KG-1GRLL*	

Fig. 6/1

EZH2 長さ : 2600bp (コ-ティグ: 90-2330)

1	AGGCAGTGGAGCCCCCGGCGCGCGCGCGCGCGCGCGGGGCGACGCCGGAARCAACG	60
61	CGAGTCGGCGCGCGGGGACGAAGAATAATCATGGGCCAGACTGGGAAGAATCTGAGAAGG M G Q T G K K S E K G	120
121	GACCAGTTTGTTGGCGGAAGCGTGTAANAATCAGAGTACATGCCACTGAGACAGCTCAAGA P V C W R K R V K S E Y M R L R Q L K R	180
181	GGTTCAGACGAGCTGATGAAGTAAAGAGTATGTTTAGITCCAATCGTCAGAAAATTTTGG F R R A D E V K S M F S S N R Q K I L E	240
241	AAAGAACGGAATCTTAAACCAAGAATGGAACAGCGAAGGATACAGCCTGTGCACATCC R T E I L N Q E W K Q R R I Q P V H I L	300
301	TGACTTCTGTGAGCTCATTGCGCGGAGTACGGAGTGTTCGGTGACCACTGACTTGGATT T S V S S L R G T R E C S V T S D L D F	360
361	TTCCAACACAAGTCATCCCATTAAAGACTCTGAATGCAGTTGCTTCAGTACCATAATGT P T Q V I P L K T L N A V A S V P I M Y	420
421	ATTCTTGGTCTCCCCCTACAGCAGAAATTTATGGTGAAGATCAAACCTGTTTACATAACA S W S P L Q Q N F M V E D E T V L H N I	480
481	TTCCITTATATGGGAGATGAAGTTTTAGATCAGGATGGTACTTTCATTGAAGAACTAATAA P Y M G D E V L D Q D G T F I E E L I K	540
541	AAAATTATGATGGGAAAGTACACGGGATAGAGAATGTGGGTTTATAAATGATGAAATTT N Y D G K V H G D R E C G F I N D E I F	600
601	TTGTGGAGTTGGTGAATGCCCTTGGTCAATATAATGATGATGACGATGATGATGATGGAG V E L V N A L G Q Y N D D D D D D D G D	660
661	ACGATCCTGAAGAAAGAGAAGAAAAAGCAGAAAGATCTGGAGGATCACCGAGATGATAAAG D P E E R E E K Q K D L E D H R D D K E	720
721	AAAGCGCCACCTCGGAAATTTCTTCTGATAAAATTTTGAAGCCATTTCTCAATGT S R P P R K F P S D K I F E A I S S M F	780
781	TTCCAGATAAGGGCACAGCAGAAGAACTAAAGGAAAAATATAAGAACTCACCGAACAGC P D K G T A E E L K E K Y K E L T E Q Q	840
841	AGCTCCCAGCCCACTTCTCTCTGAATGTACCCCCAACATAGATGGACCAATGCTAAAT L P G A L P P E C T P N I D G P N A K S	900
901	CTGTTCCAGAGAGAGCAAGCCTTACACTCCTTTTCATACGCTTTTCTGTAGCGGATGTTTA V Q R E Q S L H S F H T L F C R R C F K	960
961	AATATGACTGCTTCTACATCCTTTTCATGCAACACCCACACTTATAAGCGGAGAACAA Y D C F L H P F H A T P N T Y K R K N T	1020
1021	CAGAAACAGCTCTAGACACAAACCTTGTGGACCACAGTGTTACCAGCAATTGGAGGGAG E T A L D N K P C G P Q C Y O H L E G A	1080
1081	CAAAGGAGTTTCTGCTGCTCTCACCGCTGAGCGGATAAAGACCCCAACAAAACGTCACG K E F A A A L T A E R I K T P P K R P C	1140

Fig. 6/2

1141	GAGGCCGCAGAAGAGGACGGCTTCCCAATAACAGTAGCAGGCCAGCACCCCCACCATTA	1200
	G R R R G R L P N N S S R P S T P T I N	
1201	ATGTGCTGGAATCAAAGGATACAGACAGTGATAGGGAAGCAGGGACTGAAACGGGGGGAG	1260
	V L E S K D T D S D R E A G T E T G G E	
1261	AGAACAAATGATAAGAAGAAGAAGAAGAAGATGAACTTCGAGCTCCTCTGAAGCAA	1320
	N N D K E E E E K K D E T S S S S E A N	
1321	ATTCTCGGTGTCAAACACCAATAAAGATGAAGCCAAATATTGAACCTCCTGAGAATGTGG	1380
	S R C Q T P I K M K P N I E P P E N V E	
1381	AGTGGAGTGTGTGAAGCCTCAATGTTTAGAGTCTCATTGGCACTTACTATGACAATT	1440
	W S G A E A S M F R V L I G T Y Y D N F	
1441	TCTGTGCCATTGCTAGGTTAATTGGGACCAAAACATGTAGACAGGTGTATGAGTTTAGAG	1500
	C A I A R L I G T K T C R Q V Y E F R V	
1501	TCAAAGAATCTAGCATCATAGCTCCAGCTCCCGCTGAGGATGTGGATACTCCTCCAAGGA	1560
	K E S S I I A P A P A E D V D T P P R K	
1561	AAAAGAAGAGGAAACACCGGTTGTGGGCTGCACACTGCAGAAAGATACAGCTGAAAAAGG	1620
	K K R K H R L W A A H C R K I Q L K K D	
1621	ACGGCTCCTCTAACCATGTTTACAACATCAACCTGTGATCATCCACGGCAGCCTTGTG	1680
	G S S N H V Y N Y Q P C D H P R Q P C D	
1681	ACAGTTCGTGCCCTTGTGTGATAGCACAAAATTTTGTGAAAAGTTTGTCAATGTAGTT	1740
	S S C P C V I A Q N F C E K F C Q C S S	
1741	CAGAGTGTCAAACCGCTTTCCGGGATGCCGCTGCAAAGCACAGTGCAACACCAAGCAGT	1800
	E C Q N R F P G C R C K A Q C N T K Q C	
1801	GCCCGTGCTACCTGGCTGTCCGAGAGTGTGACCTGACCTCTGTCTTACTTGTGGAGCCG	1860
	P C Y L A V R E C D P D L C L T C G A A	
1861	CTGACCATTGGGACAGTAAAAATGTGTCTGCAAGAACTGCAGTATTACAGCGGGGCTCCA	1920
	D H W D S K N V S C K N C S I Q R G S K	
1921	AAAAGCATCTATTGCTGGCAACCATCTGACGTGGCAGGCTGGGGGATTTTATCAAAGATC	1980
	K H L L L A P S D V A G W G I F I K D P	
1981	CTGTGCAGAAAAATGAATTCTCTCAGAATACTGTGGAGAGATTATTTCTCAAGATGAAG	2040
	V Q K N E F I S E Y C G E I I S Q D E A	
2041	CTGACAGAAGAGGGAAGTGTATGATAAATACATGTGCAGCTTCTGTTCAACTTGAACA	2100
	D R R G K V Y D K Y M C S P L F N L N N	
2101	ATGATTTTGTGCTGATGCAACCCGCAAGGGTAACAAAATTCGTTTTGCAATCATTGGG	2160
	D F V V D A T R K G N K I R F A N H S V	
2161	TAAATCCAAACTGCTATGCAAAAGTTATGATGGTTAACGGTGATCACAGGATAGGTATTT	2220
	N P N C Y A K V M M V N G D H R I G I F	
2221	TTGCCAAGAGAGCCATCCAGACTGGCGAAGAGCTGTTTTTGTATTACAGATACAGCCAGG	2280
	A K R A I Q T G E E L F F D Y R Y S Q A	

【図6】

Fig. 6/3

2281	CTGATGCCCTGAAGTATGTGGCATCGAAAGAGAAATGGAAATCCCTTGACATCTGCTAC	2340
	D A L K Y V G I E R E M E I P *	
2341	CTCCTCCCCCTCCTCTGAAACAGCTGCCTTAGCTTCAGGAACCTCGAGTACTGTGGGCAA	2400
2401	TTTAGAAAAAGACATGCAGTTTGAAATTCTGAATTTGCAAAGTACTGTAAGAATAATTT	2460
2461	ATAGTAATGAGTTTAAAAATCAACTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTTTG	2520
2521	TACCAGTGAATTTTGGCAATAATGCAGTATGGTACATTTTCAACTTTGAATAAAGAATA	2580
2581	CTTGAACCTGTCAAAAAAAA	2600

【図7】

Fig. 7/1

SUV39H 長さ: 2732 bp (コ-ピ-グ: 45-1284)

1	TCGCGAGGCCGGCTAGGCCCGAATCTCGTTAGCCGTGGGAAAGATGGCGGAAAATTTAA	60
	M A E N L K	
61	AAGGCTGCAGCGTGTGTTGCAAGTCTTCTTGAATCAGCTGCAGGACCTGTGCCGCTGG	120
	G C S V C C K S S W N Q L Q D L C R L A	
121	CCAAGCTCTCCTGCCCTGCCCTCGGTATCTCTAAGAGGAACCTCTATGACTTTGAAGTCG	180
	K L S C P A L G I S K R N L Y D F E V E	
181	AGTACCTGTGGCATTACAAGAAGATCCGCGAACAGGAATATTACCTGCTGAAATGCCGTG	240
	Y L C D Y K K I R E Q E Y Y L V K W R G	
241	GATATCCAGACTCAGAGAGCACCTGGGAGCCACGGCAGAATCTCAAGTGTGTGCGTATCC	300
	Y P D S E S T W E P R Q N L K C V R I L	
301	TCAAGCAGTTCCACAAGGACTTAGAAAGGGAGCTGCTCCGGCGGCACCACCGGTCAAAGA	360
	K Q F H K D L E R E L L R R H H R S K T	
361	CCCCCGGCACCTGGACCCAGCTTGGCCAACTACCTGGTGCAGAAGGCCAAGCAGAGGC	420
	P R H L D P S L A N Y L V Q K A K Q R R	
421	GGGCGCTCCGTGCTGGGAGCAGGAGCTCAATGCCAAGCGCAGCCATCTGGGACGCATCA	480
	A L R R W E Q E L N A K R S H L G R I T	
481	CTGTAGAGAATGAGGTGGACCTGGACGGCCCTCCGCGGGCCTTCGTGTACATCAATGAGT	540
	V E N E V D L D G P P R A F V Y I N E Y	
541	ACCGTGTGTGGTGAGGGCATCACCTCAACCAGGTGGCTGTGGGCTGCCAGTCCCAGGACT	600
	R V G E G I T L N Q V A V G C E C Q D C	
601	GTCTGTGGGCACCCACTGGAGGCTGCTGCCCGGGGGCGTCACTGCACAAGTTTGCCCTACA	660
	L W A P T G G C C P G A S L H K F A Y N	
661	ATGACCAGGGCCAGGTGCGGCTTCGAGCCGGGCTGCCCATCTACGAGTGCAACTCCCGCT	720
	D Q G Q V R L R A C L P I Y E C N S R C	
721	GCCGCTGCGGCTATGACTGCCCAAATCGTGTGGTACAGAAGGGTATCCGATATGACCTCT	780
	R C G Y D C P N R V V Q K G I R Y D L C	
781	GCATCTTCCGGACGGSATGATGGGCGTGGCTGGGCGCTCCGCACCTGGAGAAGATTCCGA	840
	I F R T D D G R G W G V R T L E K I R K	
841	AGAACAGCTTCGTATGGAGTACGTGGGAGAGATCATTACCTCAGAGGAGGCAGAGCGGC	900
	N S F V M E Y V G E I I T S E E A E R R	
901	GGGGCCAGATCTACGACCGTACGGGCGCCACCTACCTCTTTGACCTGGACTACGTGGAGG	960
	G Q I Y D R Q G A T Y L F D L D Y V E D	
961	ACGTGTACACCGTGGATGCCGCCTACTATGGCAACATCTCCCACTTTGTCAACCACAGTT	1020
	V Y T V D A A Y Y G N I S H F V N H S C	
1021	GTGACCCCAACCTGCAGGTGTACAACGTCTTCATAGACAACCTTGACGAGCGGCTCCCCC	1080
	D P N L Q V Y N V F I D N L D E R L P R	
1081	GCATCGCTTTCTTTGCCACAAGAACCATCCGGGCGAGGAGGAGCTCACCTTTGATTACA	1140
	I A F F A T R T I R A G E E L T F D Y N	

Fig. 7/2

1141	ACATGCAAGTGGACCCCGTGGACATGGAGAGCACCCGCATGGACTCCAACCTTGGCCTGG	1200
	M Q V D P V D M E S T R M D S N F G L A	
1201	CTGGGCTCCCTGGCTCCCCAAGAAGCGGGTCCGTATTGAATGCAAGTGTGGGACTGAGT	1260
	G L P G S P K K R V R I E C K C G T E S	
1261	CCTGCCGCAAATACCTCTTCTAGCCCTTAGAAGTCTGAGGCCAGACTGACTGAGGGGGCC	1320
	C R K Y L F *	
1321	TGAAGCTACATGCACCTCCCCCACTGCTGCCCTCCTGTCGAGAATGACTGCCAGGGCCTC	1380
1381	GCCTGCCTCCACCTGCCCCCACTGCTCCTAOCCTGCTCTACGTTGAGGGCTGTGGCCGTG	1440
1441	GTGAGGACCGACTCCAGGAGTCCCCCTTCCCTGTCCCAGCCCCATCTGTGGGTGCACTT	1500
1501	ACAAACCCCCACCCACCTTCAGAAATAGTTTTTCAACATCAAGACTCTCTGTGCTGGGA	1560
1561	TTGATGGCCTATTAAAGGAGGTCCAAGGGGTGAGTCCCAACCCAGCCCCAGAAATATATTG	1620
1621	TTTTTGACCTGCTTCTGCCTGGAGATTGAGGGGTCTGCTGCAGGCCTCCTCCCTGCTGC	1680
1681	CCCAAAGGTATGGGGAAGCAACCCAGAGCAGGCAGACATCAGAGGCCAGAGTGCCTAGC	1740
1741	CCGACATGAAGCTGGTTCCCCAACCCAGAACTTTGTACTAGTGAAAGAAAGGGGTCCC	1800
1801	TGGCCTACGGGCTGAGGCTGGTTTCTGCTCGTGTCTACAGTCTGCGTAGTGTGGCCCT	1860
1861	AAGAGCTGTAGGGTCTCTTCTCAGGGCTGCATATCTGAGAAGTGGATGCCACATGCCA	1920
1921	CTGGAAGGGAAGTGGGTGTCCATGGGCCACTGAGCAGTGAGAGGAAGGCAGTGACAGCT	1980
1981	GGCCAGCCCTGAGAGTAGGCTGGGACCAAGCTCTGCCTTCACAGTGCAGTGAAGGTACCT	2040
2041	AGGGCTCTTGGGAGCTCTGCGGTGCTAGGGGCCCTGACCTGGGGTGTGATGACCGTGA	2100
2101	CACCACTCAGAGCTGGAACCAAGATCTAGATAGTCCGTAGATAGCACTTAGGACAAGAAT	2160
2161	GTGCATTGATGGGGTGGTGTGATGAGGTGCCAGGCACTAGGTAGAGCACCCTGGTCCACGTGG	2220
2221	ATTGTCTCAGCGAAGCCTTGAAAACACGGAGGTGGATGCCAGGAAGGGCCCATGTGCC	2280
2281	AGAAGGCAAGTACAGGCCAAGAATTGGGGTGGGGGAGATGGCTTCCCCACTATGGGAT	2340
2341	GACGAGGCGAGAGGGAAGCCCTTGCTGCCTGCCATTCCCAGACCCAGCCCTTTGTGCTC	2400
2401	ACCCTGGTTCCACTGGTCTCAAAAGTCACCTGCCTACAAATGTACAAAAGGCGAAGGTTT	2460
2461	TGATGGCTGCCCTTGCTCCTTGCTCCCCACCCCTGTGAGGACTTCTCTAGGAAGTCTT	2520
2521	CCTGACTACCTGTGCCAGAGTGCCCTACATGAGACTGTATGCCCTGCTATCAGATGCC	2580
2581	AGATCTATGTGTCTGTCTGTGTCCATCCCGCCGGCCCCCAGACTAACCTCCAGGCAT	2640
2641	GCACTGAATCTGGTTCTCTCTGTGTACACCCCTCAACCCATGCAGCCTGGAGTGGGCAT	2700
2701	CAATAAAATGAACTGTCGACTGAAAAA	2732

1 AATCATGATTAAATATCTGGTATCATTITAGGCCCTCTC 489

INTERNATIONAL SEARCH REPORT

Inter national Application No
PCT/EP 96/01818

A. CLASSIFICATION OF SUBJECT MATTER		
IPC 6	C12N15/12 A01K67/00	C07K14/47 C12N1/21 C07K16/18 C12N5/10 C12N15/11 G01N33/50
According to International Patent Classification (IPC) or to both national classification and IPC.		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols)		
IPC 6 C12N C07K G01N A01K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	DATABASE EMBL entry HS529651, accession number U52965, 26 April 1996 HOBERT O ET AL: "Interaction of Vav with ENX-q, a putative transcriptional regulator of homeobox gene expression" XP002017786 see sequence	1-12,16
T	& MOLECULAR AND CELLULAR BIOLOGY, vol. 16, no. 6, June 1996, WASHINGTON US, pages 3066-3073, XP000609032 HOBERT O ET AL: "Interaction of Vav with ENX-1, a putative transcriptional regulator of homeobox gene expression" see figure 1 --- -/--	1-12,16
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "Z" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
7 November 1996		15. 11. 96
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016		Authorized officer Espen, J

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/EP 96/01818

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL entry HSMG44A, accession number L08238, January 1993 GERAGHTY MT: "Human MG44 mRNA, 5' end" XP002017787 cited in the application see sequence & GENOMICS, vol. 16, 1993, pages 440-446, XP000609031 GERAGHTY ET AL: "The isolation of cDNAs from OATL1 at Xp11.2 using a 480-kb YAC" cited in the application see the whole document ----	1-12,16
A		
X	DATABASE EMBL entry HS18003, accession number U18003, December 1994 OSTERMEYER EA: "Human chromosome 17q21 mRNA clone B117" XP002017911 siehe Sequenz & GENOMICS, vol. 25, January 1995, pages 256-263, XP000609030 FRIEDMAN LS: "22 genes from chromosome 17q21: cloning, sequencing and characterization of mutations in breast cancer families and tumors" see the whole document ----	1-12,16
X		1-12,16
X	MOLECULAR AND CELLULAR BIOLOGY, vol. 13, 1993, WASHINGTON US, pages 6357-6366, XP000608922 JONES RS ET AL: "The Drosophila polycomb-group gene enhancer of zeste contains a region with sequence similarity to trithorax" cited in the application see figure 3 -----	1,2
X	EMBO JOURNAL, vol. 13, 1994, EYNSHAM, OXFORD GB, pages 3822-3831, XP002017785 TSCHERSCH B ET AL: "The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes" cited in the application see page 5 -----	1,2

Form PCT/ISA/210 (continuation of Form PCT/ISA/210) (July 1992)

フロントページの続き

(51) Int. Cl. ⁷	識別記号	F I	テーマコード (参考)
C 1 2 N 5/10		C 1 2 Q 1/68	A
C 1 2 Q 1/68		G 0 1 N 33/50	Z
G 0 1 N 33/50		C 1 2 N 5/00	B